REPORT RESUMES

ED 015 010 PS 000 219

A STUDY OF THE FULL-YEAR 1966 HEAD START PROGRAMS.

BY- CORT, H. RUSSELL, JR. AND OTHERS

PLANNING RESEARCH CORP., WASHINGTON. D.C.

REPORT NUMBER PRC-R-886 PUB DATE 31 JUL 67

REPORT NUMBER OEO-1308

EDRS PRICE MF-$1.25 HC-$10.84 269P.

DESCRIPTORS- *PRESCHOOL LEARNING, *PRESCHOOL EVALUATION, *PERFORMANCE TESTS, *ACHIEVEMENT TESTS, RACIAL CHARACTERISTICS, STUDENT CHARACTERISTICS, PHYSICAL CHARACTERISTICS, TEACHER CHARACTERISTICS, PROGRAM EVALUATION, PARENT PARTICIPATION, FAMILY STRUCTURE, FAMILY CHARACTERISTICS, SOCIOECONOMIC BACKGROUND, *PROGRAM EFFECTIVENESS, PSI, PPVT, VSMS, HEAD START, BI, DAP,

AS PART OF THE EVALUATION OF EFFECTIVENESS OF FULL-YEAR HEAD START PROGRAMS, CHILDREN FROM A NATIONWIDE SAMPLE OF CENTERS WERE TESTED WITH FIVE INSTRUMENTS (PEABODY PICTURE VOCABULARY TEST, PRE-SCHOOL INVENTORY, VINELAND SOCIAL MATURITY SCALE, DRAW-A-PERSON, AND BEHAVIOR INVENTORY). CENTERS WERE SELECTED TO BE REPRESENTATIVE OF PROGRAMS OF DIFFERENT LENGTHS. POST-TESTS WERE USED TO EXAMINE THE QUESTION OF WHETHER THE LENGTH OF THE PROGRAM AFFECTS THE PERFORMANCE OF THE CHILDREN. THERE WAS NO RELIABLE EVIDENCE OF A SYSTEMATIC RELATIONSHIP BETWEEN LENGTH OF PARTICIPATION IN A PROGRAM AND LEVEL OF PERFORMANCE OR DEVELOPMENT. FACTORS AFFECTING THE INTERPRETATION OF RESULTS ARE DISCUSSED IN THIS REPORT. (DESCRIPTIVE STATISTICS ON THE TEST SCORES, CHARACTERISTICS OF PARENTS, FAMILIES, AND STAFF MEMBERS ARE PRESENTED. INTER-TEST CORRELATIONAL DATA ARE REPORTED.) A SUMMARY OF THIS REPORT IS ALSO AVAILABLE AS A SEPARATE DOCUMENT. (LG)

ED015010

A STUDY OF THE FULL-YEAR 1966 HEAD START PROGRAMS

PRC R-886

31 July 1967

Prepared for
Office of Economic Opportunity
Project Head Start Division of Research and Evaluation

PS000219

PLANNING RESEARCH CORPORATION
LOS ANGELES, CALIFORNIA WASHINGTON, D.C.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

A STUDY OF THE FULL-YEAR 1966 HEAD START PROGRAMS

PRC R-886

31 July 1967

Prepared for
Office of Economic Opportunity
Project Head Start
Division of Research and Evaluation
Under Contract OEO 1308

By
H. Russell Cort, Jr.
William D. Commins, Jr.
Naomi H. Henderson
Margaret A. Mattis
Ruth Ann O'Keefe
Reginald C. Orem

PLANNING RESEARCH CORPORATION
Universal Building North • Suite 1030 • 1875 Connecticut Avenue N.W. • Washington, D.C. 20009

PS000219

FOREWORD

This is a report on the results of a study of the 1966 full-year Head Start program, prepared by Planning Research Corporation (PRC) in fulfillment of Contract OEO 1308. The study was part of an assessment program conducted by the Research and Evaluation Division of Project Head Start, under the direction of Dr. Edmund W. Gordon.

The report is the result of the efforts of many people. PRC would like to express its appreciation to the OEO staff members in the national Washington office and in the many local programs which participated in the study, for their assistance and cooperation. The U.S. Bureau of the Census had responsibility for the data processing, and special thanks go to Mr. Leonard Goldberg, Mr. William T. Allsbrook, and Mr. Robert Hanson for their assistance in many aspects of this task.

In addition, the following individuals contributed special effort and time to the project:

- PRC Staff Members:
    - Miss Leslie M. Appleton
    - Miss Patricia Eller
    - Mr. Carl Morgan
    - Mrs. Janet Duca Norton
    - Mrs. Susan May Pothier
    - Mr. Donald Singman
    - Mrs. Susan Starr
- PRC Publications and Graphics Departments
- Mrs. Louise Crenshaw, Bryn Mawr College
- Miss Marilyn Danner
- Miss Carol J. Kline
- Mrs. Betty Krone, Office of Economic Opportunity
- Mrs. Virginia Raney, Office of Economic Opportunity
- Mr. Leonard Oliver, Center of Adult Education, University of Maryland
- Dr. Philip Porter, Harvard Medical School

PRC also wishes to express appreciation to the consultants who participated in the tester-training program: Dr. Charles Dailey and Dr. Charles Shaffer, The American University; Mr. Karl Banks, D.C. Junior Village; and Dr. Lois-ellin Datta, National Institute of Mental Health. To Dr. Datta go special thanks for her interest, involvement, and suggestions throughout the project.

Finally, PRC gratefully acknowledges the Head Start teachers, and the individuals who actually tested the children. The testers, who are named individually in Appendix C, often served "beyond the call of duty;" their cooperation and diligence provided much of the information from which assessments of the programs were made.

The project was completed under the general supervision of Dr. Allen R. Ferguson, Deputy Manager of PRC's Systems Economics Division, and Dr. Norman H. Jones, Manager of PRC's Economics Department. Dr. H. Russell Cort, Jr. was the Project Manager.

ABSTRACT

As part of the evaluation of the effectiveness of full-year Head Start programs, children from a nationwide sample of centers were tested or rated with five different instruments. The centers were selected to be representative of programs of different lengths. Testing was performed at the end of the programs. Analyses were made to test the hypothesis that the length of the program affects the performance of the children. There was no reliable evidence of a systematic relationship between length of participation in a program and level of performance or development as measured by the Peabody Picture Vocabulary Test (PPVT), the Pre-School Inventory (PSI), or the Vineland Social Maturity Scale (VSMS). There was a significant decrease in scores for intermediate-length program children on the Behavior Inventory (BI); however, that result did not appear to be directly related to the effects of the program on children. Factors affecting the interpretation of results are discussed in this report.

Descriptive statistics on the test scores, characteristics of parents and families, and characteristics of Child Development Center (CDC) staff members are presented. Intertest correlational data are also reported. Data and analysis of Draw-A-Person (DAP) tests are also presented.

TABLE OF CONTENTS

Page

I. INTRODUCTION 1
  A. Background 1
  B. Report Organization 1
II. STUDY OBJECTIVES 3
III. APPROACH 5
  A. General Approach 5
  B. Tests 6
  C. Experimental Design 8
    1. The Universe 8
    2. Sample Size 9
    3. Mechanics of Drawing Centers 10
    4. Replacement of Centers 12
    5. Mechanics of Drawing Children 12
  D. Primary Data Collection 13
    1. Selection and Training of Testers 13
    2. Individual Tests 14
    3. Rating Scales 15
    4. Testing Supervision 15
    5. Spanish-Speaking Children 15
  E. Supplementary Data Collection 16
IV. RESULTS 17
  A. Composition of Resultant Samples 17
    1. Children 17
    2. Staff Members - Distributions by Characteristics 26
    3. Families - Distributions by Characteristics 33
  B. Test Results 42
    1. Distributions of Raw Scores and Unadjusted Means by Short-, Medium-, and Long-Term Centers 42
    2. Statistical Analyses 92

TABLE OF CONTENTS
(Continued)

Page

V. DISCUSSION 125

A. The Main Hypothesis. 125

B. Alternative Hypotheses 126

1. Non-Comparability of Samples on Essential Uncontrolled Variables 127

2. Inadequate Sample Sizes 129

3. Non-Uniform Treatments and Effects Among or Within Samples 131

4. Inappropriate Instruments or Tests 135

5. Effects of Head Start Which May Not Be Immediately Noticeable. 155

C. Exceptions 156

1. The Behavior Inventory 156

2. Children Tested in Spanish. 158

VI. CONCLUSIONS. 161

APPENDIX A SAMPLES A-1

APPENDIX B STATISTICAL MODELS AND ANALYSES . . . B-1

APPENDIX C TESTERS: SELECTION, TRAINING, AND SUPERVISION. C-1

APPENDIX D TESTS D-1

APPENDIX E EXAMINATIONS OF SELECTED CENTERS. . E-1

APPENDIX F THE DRAW-A-PERSON TEST RESULTS . . . F-1

LIST OF EXHIBITS

Page

1. Locations of Short-, Medium-, and Long-Term Centers 19

2. Sample Compositions in Numbers and Percentages of All Children Tested, by Sex, Race, and Age 21

3. Sample Compositions in Numbers and Percentages of "Eligible" Children Only, by Sex, Race, and Age . . 22

4. Summary of Eligible Children, by Program Duration, Sex, Age, and Language Used in Testing 24

5. Number of Untestable and Untested (Absent) Children, by Sex, Race, and Age 25

6. Percentage of Untestable Children by Sex, Race, and Age 27

7. Percentage of Staff Members by Staff Position, for S, M, and L Centers 28

8. Age Distributions of Staff Members by Staff Position, for Short-, Medium-, and Long-Term Centers 30

9. Cultural and Ethnic Characteristics of Staff Members by Staff Position, for Short-, Medium-, and Long-Term Centers 31

10. Percentage of Paid Versus Volunteer Staff Members by Staff Position, for Short-, Medium-, and Long-Term Centers 32

11. Number of Years of Experience With Pre-School Children by Staff Position, for Short-, Medium-, and Long-Term Centers 34

12. Number of Years of Experience With Children From Conditions of Poverty, by Staff Position, for Short-, Medium-, and Long-Term Centers 35

13. Percentage of Family Response Based on Sample Children. 37

LIST OF EXHIBITS
(Continued)

Page

14. Siblings and Family Structural Characteristics 37

15. Mean Family Size for Three Subgroups of Children Tested 39

16. Percentage of Families of Sample Children With Father Living at Home 39

17. Level of Education Completed by Parents of Children in Short-, Medium-, and Long-Term Centers 40

18. Annual Income of Families With Children in Short-, Medium-, and Long-Term Centers 41

19A. Distribution of PPVT Raw Scores for Children in Short-Term Centers, by Sex, Race, and Age 43

19B. Distribution of PPVT Raw Scores for Children in Medium-Term Centers, by Sex, Race, and Age 45

19C. Distribution of PPVT Raw Scores for Children in Long-Term Centers, by Sex, Race, and Age 47

20A. Distribution of PSI Raw Scores for Children in Short-Term Centers, by Sex, Race, and Age 51

20B. Distribution of PSI Raw Scores for Children in Medium-Term Centers, by Sex, Race, and Age 53

20C. Distribution of PSI Raw Scores for Children in Long-Term Centers, by Sex, Race, and Age 55

21A. Distribution of BI Raw Scores for Children in Short-Term Centers, by Sex, Race, and Age 59

21B. Distribution of BI Raw Scores for Children in Medium-Term Centers, by Sex, Race, and Age 61

21C. Distribution of BI Raw Scores for Children in Long-Term Centers, by Sex, Race, and Age 63

22A. Distribution of VSMS Raw Scores for Children in Short-Term Centers, by Sex, Race, and Age 67

LIST OF EXHIBITS
(Continued)

Page

22B. Distribution of VSMS Raw Scores for Children in Medium-Term Centers, by Sex, Race, and Age 69

22C. Distribution of VSMS Raw Scores for Children in Long-Term Centers, by Sex, Race, and Age....... 71

23. PPVT Summary of Raw Score Means and Sample Sizes for Short-, Medium-, and Long-Term Eligible Children, by Language Used in Testing, Race, and Age 76

24. PSI Summary of Raw Score Means and Sample Sizes for Short-, Medium-, and Long-Term Eligible Children, by Language Used in Testing, Race, and Age 77

25. BI Summary of Raw Score Means and Sample Sizes for Short-, Medium-, and Long-Term Eligible Children, by Language Used in Testing, Race, and Age 78

26. VSMS Summary of Raw Score Means and Sample Sizes for Short-, Medium-, and Long-Term Eligible Children, by Language Used in Testing, Race, and Age 79

27. PPVT Summary of Unadjusted Raw Score Means, Standard Errors of Means, Medians, and Sample Sizes for Eligible Children in Short-, Medium-, and Long-Term Centers 80

28. PSI Summary of Unadjusted Raw Score Means, Standard Errors of Means, Medians, and Sample Sizes for Eligible Children in Short-, Medium-, and Long-Term Centers 81

29. VSMS Summary of Unadjusted Raw Score Means, Standard Errors of Means, Medians, and Sample Sizes for Eligible Children in Short-, Medium-, and Long-Term Centers 82

30. BI Summary of Unadjusted Raw Score Means, Standard Errors of Means, Medians, and Sample Sizes for Eligible Children in Short-, Medium-, and Long-Term Centers 83

LIST OF EXHIBITS
(Continued)

Page

31. PSI Subtest Mean Raw Scores for All Children, by Age and Race 84

32. PSI Subtest Mean Raw Scores for Eligible Children, by Age and Race. 85

33. PSI Subtest Mean Raw Scores Expressed as Percent of Total Possible Scores for All Children. 86

34. PSI Subtest Mean Raw Scores Expressed as Percent of Total Possible Scores for Eligible Children . . 87

35. Percentage of Eligible Children Passing PSI Items, by Age and Program Type 88

36. Poverty as a Function of Household Size and Income 94

37. Total Sample Parameters for Dependent Variables. . . 95

38. Covariate Means and Standard Deviations 95

39. Covariance Adjustments to Test Raw Score Means by Covariate and Duration Level (S, M, L). 97

40. Multiple Regression Parameters for Each Test. 100

41. Independent Variables With Significant Beta Coefficients, by Test 101

42. Subgroups Used for By-Child Covariance Analyses. . . 105

43. Independent Variable and Unadjusted and Adjusted Dependent Variable Means by Group and Program Duration. 106

44. Subgroup Covariance Increments 107

45. Highlights of Results of Stepwise Regression Analyses by Test and Group. 115

46. Means and Standard Deviations of Independent Variables by Group 118

LIST OF EXHIBITS
(Continued)

Page

47. Intertest Product-Moment Correlation Coefficients From By-Center Multivariate Covariance Analysis. . . . 119

48. Intertest Residual Correlations 119

49. Partial Correlation Coefficients of the Independent Variables With Each of the Four Tests 120

50. Residual Correlation Coefficients Estimates and Product-Moment Correlation Coefficients for PSI Subtests and Other Tests for Group 1 121

51. Correlation of Independent and Dependent Variables for Group 1 123

52. Simultaneous Ninety-Five Percent Confidence Intervals for Differences Between Adjusted Treatment Means 130

53. Mean Behavior Inventory Scores for Hypothetical Children. 139

54. Results of Analysis of Variance of Test Scores, by Tester and Age of Children Tested 141

55. Correlations of Center Opinions and Tests 144

56. Percentages of Children Rated Negatively on Selected Test Behavior Variables 147

57. Association of Low Ratings on Selected PPVT Test Behavior Scales With PPVT IQ 148

58. KR-20 Coefficients for PSI Subtests 154

59. Partial Correlations for Spanish-Speaking Group 5 . . . 159

A-1. List of Eligible Short-Term Census Sample Centers From Which the PRC Subsample Was Drawn. A-2

A-2. List of Eligible Medium-Term Census Sample Centers From Which the PRC Subsample Was Drawn A-5

LIST OF EXHIBITS
(Continued)

Page

A-3. List of Eligible Long-Term Census Sample Centers From Which the PRC Subsample Was Drawn A-7

A-4. List of Centers Tested by PRC A-9

B-1. Power of PPVT F-Test B-3

B-2. Combination of the Lambda's B-5

B-3. PPVT Residuals Scatter Plot B-7

B-4. Range of Residuals B-8

C-1. Head Start Testers C-7

C-2. Schedule for First Head Start Tester Training Program, May 12-15, 1966 . . . . C-8

C-3. Schedule for Second Head Start Tester Training Program, June 8-11, 1966 . . . . C-10

C-4. Extract from PRC's Tester's Handbook C-13

D-1. Mean Time in Minutes Required to Administer the PPVT and PSI to Three-, Four-, Five-, and Six-Year-Old Children D-18

D-2. English Version of Family Information Form (CAP-HS Form 46) D-20

D-3. Spanish Version of Family Information Form (CAP-HS Form 46a) D-25

D-4. Behavior Inventory (CAP-HS Form 37) D-30

E-1. Selected Characteristics of Those Centers With the Top Ten and the Bottom Ten Sums of Raw Score Means E-4

E-2. Description of Centers With the Ten Highest PPVT IQ Means E-8

LIST OF EXHIBITS
(Continued)

Page

E-3. Description of Centers With the Ten Lowest PPVT IQ Means E-9

E-4. Selected Test Data of Those Centers With the Top Ten and the Bottom Ten Sums of Raw Score Means E-12

E-5. Anecdotal Material From Testers' Reports of Those Centers With the Top Ten Sums of Raw Score Means E-15

E-6. Anecdotal Material From Testers' Reports of Those Centers With the Bottom Ten Sums of Raw Score Means E-16

E-7. Anecdotal Material From Testers' Reports of Those Centers With the Top Ten PPVT IQ Means E-19

E-8. Anecdotal Material From Testers' Reports of Those Centers With the Bottom Ten PPVT IQ Means E-20

E-9. Top Ten and Bottom Ten Centers Based on Residuals E-21

E-10. Analysis of Classroom Effects Within a Sample of Centers E-24

F-1. Interscorer Reliability for Person Drawings F-3

F-2. Distribution of DAP Raw Scores for Children in Short-Term Centers, by Sex, Race, and Age F-4

F-3. Distribution of DAP Raw Scores for Children in Medium-Term Centers, by Sex, Race, and Age F-6

F-4. Distribution of DAP Raw Scores for Children in Long-Term Centers, by Sex, Race, and Age F-8

F-5. Intertest Correlations for Total and Subgroups: PPVT Raw Score, PPVT IQ, DAP Raw Score, DAP IQ, and PSI Raw Score F-12

LIST OF EXHIBITS
(Continued)

Page

F-6. Intertest Correlations for Three-, Four-, Five-, and Six-Year-Old PPVT Raw Score, PPVT IQ, DAP Raw Score, DAP IQ, and PSI Raw Score F-15

F-7. Raw Score IQ and SQ Means and Standard Deviations F-16

I. INTRODUCTION

A. Background

The study reported here was undertaken as part of a national program of evaluation of the 635 1966 full-year Head Start programs. When the study was initiated, a major data collection effort had already been designed and implemented by the U.S. Bureau of the Census. It was the task of the Census Bureau to obtain from a national sample of Child Development Centers (CDC's), children, and parents, standardized information about:

- The age, sex, and race characteristics of the children.
- The medical and dental history and status of the children.
- The characteristics of staff members and workers in the sample CDC's.
- The evaluation of the individual programs by staff members and workers.
- The participation of parents in the local programs.
- The structure and other socio-economic characteristics of the families of the children.

The data collected by the Census Bureau have provided, for the most part, important descriptive information about the characteristics of various elements of the full-year programs. They will not (and were not intended to) provide measures of the performance or achievement of the children served. It was the purpose of the present study to fill this need partially and thus to supply an important dimension in the evaluation of program effectiveness at the national level.

B. Report Organization

This report is organized primarily in two parts. Sections II through VI describe the study and present and discuss results. Appendixes A through F provide amplifying and supplementary material. In Section II, the objectives of the study are stated. Section III provides a description of the study design, the tests used and measures obtained, the experimental design, and the data collection procedures. Section IV presents the results, Section V discusses these results, and Section VI lists conclusions. The appendixes provide details on the sample centers and sample universes (Appendix A), the statistical models (Appendix B), the training and supervision of testers (Appendix C), the tests (Appendix D), a subsidiary analysis of high- and low-scoring centers (Appendix E), and the Draw-A-Person test results (Appendix F).

II. STUDY OBJECTIVES

The primary objective of this study was to evaluate the effectiveness of 1965-66 full-year Head Start programs for culturally deprived pre-school children. Unlike the summer Head Start programs, which ran for approximately 8 weeks, the full-year programs during the first full year varied enormously in starting times, ending times, and length. Some programs commenced in the fall of 1965. Others were still coming into being in May 1966, when this study was initiated. Some programs ended in March or April of 1966; others did not end until January or February of 1967. Some were conducted for over 40 weeks, and others for less than 10. This very diversity offered an opportunity for an evaluation of the effectiveness of the programs that was quite impossible with the short, fixed-length summer Head Start programs. From the point of view of program planning and resource allocation, it would be extremely valuable to know something about the relationship between length (as well as kind) of program and output, performance, or effectiveness.

This study undertook to accomplish the objective stated above by testing the hypothesis that the length of a Head Start program affects the level of performance, achievement, or behavior of Head Start children.

A secondary objective of this study was to provide further descriptive data on the performance and behavioral characteristics of Head Start children in cognitive, social, and emotional areas of development. As part of the accomplishment of both objectives, Planning Research Corporation (PRC) also wished to examine some possible relationships between workers and programs on the one hand, and the performance of the children on the other.

III. APPROACH

A. General Approach

The general approach of the study was to administer tests to samples of children in programs of different durations, and to obtain ratings of the behavior and abilities of the children from their teachers in the selected Child Development Centers. The principal independent variable was the length of the individual program; the dependent variables were the performances or ratings of the children as measured by scores on the tests or scales. While there was generally a strong correlation between the length of the program and the length of time that a given child had been in the program, the two variables are not synonymous. Consequently, two main types of analyses had to be made, based on these two separate duration or exposure variables (see Section IV).

The major methodological constraint on the general study design resulted from the timing of the study. The study was initiated in May 1966, by which time pre-testing of children was impossible. Consequently, the overall approach was based on end-testing of samples of children in sets of programs of different durations. Thus, the design was essentially that of Campbell and Stanley's Design 6,[1] except that of course prior assignment of children to treatment levels was not randomized. Three main levels or durations of program were considered. These levels are labeled Short-term, Medium-term, and Long-term (S, M, and L) and are defined precisely in subsection III.C below. The sampling universe of programs consisted of a subset of CDC's in the Census Bureau's nationwide sample. The rules for defining and identifying this subset are also discussed below. Sample sizes, both for number of children and number of centers, were based on estimate variances for the Peabody Picture Vocabulary Test (PPVT) from data obtained in the 1965 summer Head Start research and evaluation program.[1]

---

[1] Campbell, Donald T. and Julian C. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching." Chapter 5 in Gage, N. L. (ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

B. Tests

The primary instruments used in the testing program were the Peabody Picture Vocabulary Test (PPVT) Form B, the revised Pre-School Inventory (PSI), the Operation Head Start Behavior Inventory (BI), and the Vineland Social Maturity Scale (VSMS). A Draw-A-Person (DAP) test was also administered.

The PPVT is an individual test of verbal ability which does not require a verbal response; for example, the tester shows a child a page containing four pictures and asks, "Show me 'table'." There are 150 possible pictures which the child may identify; the "raw score" is essentially the number of correctly identified pictures and can be converted into mental age or intelligence quotient equivalents if desired. Form B of the test was used throughout this study.[2]

The revised PSI is an individually administered 85-item measure of school readiness originally developed for the 1965 summer Head Start program by Dr. Bettye Caldwell and Mr. Donald Soule of the New York State University at Syracuse. The test provides a total score, as well as four subtest scores of separate factors related to school readiness: personal-social responsiveness, associative vocabulary, numerical concept activation, and sensory concept activation. The tester administers all items to the child; the highest possible score is 90, as five items have a score value of 2 points, while all other items have a score value of 1 point. The total raw score can be broken down into scores for each of the four factors or subtests.

---

[1] Planning Research Corporation, PRC R-795, Results of the Summer 1965 Head Start Program, Volumes I and II, H.R. Cort, Jr. et al, May 1966 (Unclassified).

[2] Dunn, Lloyd M., Expanded Manual, Peabody Picture Vocabulary Test. American Guidance Service, Inc., 720 Washington Ave. S.E., Minneapolis, Minnesota, 1965.

The BI is an instrument originally developed by Dr. Edward Zigler for the 1965 summer Head Start program. It is a set of 50 rating scales which are intended to obtain information on a number of behavioral aspects of children. Twenty-five of the scales are intended to tap positive behavioral characteristics (such as "Is usually carefree; rarely becomes apprehensive or frightened") and 25 are intended to tap negative characteristics (such as "Has little respect for the rights of other children; refuses to wait his turn, usurps toys other children are playing with," etc.). The teacher, or someone else who knows the child well, rates each child in her class on each of the 50 items or scales. The BI can provide an overall behavior-adjustment score and/or a separate adjustment score for each of nine behavioral categories (sociability-cooperation-politeness; independence-dependence; curiosity-enthusiasm-exploration-creativity; persistence; emotionality; self-confidence; jealousy-attention-seeking; achievement; and leadership).

The VSMS is an interview schedule which is given to someone who knows the child well; it provides an indication of the child's social development, maturity, and independence. First developed by Dr. Edgar A. Doll in 1935, the VSMS has a 1965 edition which was used throughout this testing program. The scale attempts to evaluate the child in eight different areas: self-help, general; self-help, eating; self-help, dressing; locomotion; occupation; communication; self-direction; and socialization.[1] The VSMS provides a total score and conversions to Social Age (SA) and Social Quotient (SQ) can be made.

The DAP (or Draw-A-Man, as originally conceived by Dr. Florence Goodenough in 1926) is a simple test of intelligence in which the tester asks the child to draw a person. The drawing is later scored against a detailed list of criteria and total points for each drawing are converted into standard scores (IQ's) or mental age equivalents. In this study the children's drawings were scored according to the criteria delineated by Harris[1] for male and female drawings. See Appendix F for further details concerning administration of the DAP.

---

[1] Doll, Edgar A., Vineland Social Maturity Scale, Condensed Manual for Directions, 1965 Edition. American Guidance Service, Inc., 720 Washington Ave. S.E., Minneapolis, Minnesota, 1965.

The results of the first four tests are included in the body of the report and in the main statistical analyses described below; DAP results are reported in Appendix F.

C. Experimental Design

As stated earlier, the overall design of the study was an end-testing of children in full-year Head Start programs, with the length of the program used as the main experimental variable and test scores of children used as the main dependent variables. The study was designed essentially for a straightforward one-way analysis of variance. Details of the sampling designs and procedures are described below.

1. The Universe

Because the design variable was the length of the program, it is necessary to define starting dates, ending dates, and weeks. To this end, Friday is the last day of the week and the week of 14 to 20 May 1966 is week zero. (There are also some negative weeks.)

The Census Bureau supplied PRC with early estimates of the starting dates and the closing dates for (potentially) every full-year Head Start center. These early estimates are the dates used in defining the universe. The universe comprises all of the children in certain kinds of centers. The rules for including or excluding centers are discussed below.

If a center's starting date was in week j and its ending date in week i, then the center is said to have an $i - j$ week program. Centers ending before week 1 and after week 12 are not in the universe. Of the remaining centers, only those are in the universe whose programs are weeks:

---

[1] Harris, Dale B., Children's Drawings as Measures of Intellectual Maturity. New York: Harcourt, Brace, and World, 1963.

11, 12, 13, 14, 15 (S)
19, 20, 21, 22, 23 (M)
27, 28, 29, 30, 31, 32, ... (L)

The three groupings define a partition of the universe into three subuniverses called S, M, and L. Knowledge of the universe was derived primarily from the Census Bureau sample; therefore, further discussion must await the introduction of this sample under subsection III. C. 3.

2. Sample Size

If σ_A^2 is taken to be the variance between centers and σ_e^2 the variance within centers, then the variance of the estimate of the mean of a subuniverse is (approximately)

$$\frac{\sigma_A^2}{m} + \frac{\sigma_e^2}{mn} \qquad (1)$$

where n is the number of children sampled from each of m centers. Expression (1) emphasized the importance of including as many centers as possible in the sample; in fact, the inadequacy of early suggestions for the number of centers was argued with a series of sums, the terminal one of which is reproduced below.

For the two variances, variance component estimates are taken as values from the summer 1965 analysis of PPVT scores--i.e., set

$$\sigma_A^2 = 60.1 \qquad (2)$$

and

$$\sigma_e^2 = 130.9. \qquad (3)$$

Then, the variance of the estimate of one of the treatment (S, M, or L) means for 12 children sampled from each of 24 centers is 2.958. Thus, the noncentrality parameter for the chi-squared test of equality of treatment means if, for example, the actual means are 36, 39, and

43, is 8.26, which yields a power of 0.73. Thus, mean differences on this order should show up quite well.

The power of the design against monotonic mean increases in program length could have been increased by concentrating centers more on the S and L ends, but the risk of observing an inexplicable M average would have been too great.

Some gain in the power to distinguish among S, M, and L will result from using four scores, even though they are correlated, rather than just the PPVT.

3. Mechanics of Drawing Centers

The Census Bureau drew a sample of centers from a universe which is larger than but includes PRC's. When deletions were made from the Census Bureau sample of those centers which were in the Census Bureau's universe but not in PRC's, a second Census Bureau sample--one from PRC's universe--was constructed, as is shown in the first three exhibits of Appendix A. The PRC sample is a subsample of the second Census Bureau sample. It should be noted that PRC had to work with an early list that listed 428 centers in the Census Bureau sample; the Census Bureau sample contains somewhat more than 428 centers.

In drawing its sample, the Census Bureau used an early estimate of the number c of classes in the center. The Census Bureau provided PRC with these c's and chose centers with probability

$$\begin{array}{ll} 1/4 & \text{for } c \le 5 \\ c/20 & \text{for } 6 \le c \le 19 \\ 1 & \text{for } c \ge 20. \end{array}$$

PRC subsampled centers with probability

$$c/T_j \text{ for } c \leq 5$$

$$5/T_j \text{ for } 6 \leq c \leq 19$$

$$c/4T_j \text{ for } c \geq 20.$$

As a result, centers entered the PRC sample with probability

$$c/4T_j$$

where

$$T_S = 155/24$$

$$T_M = 99/24$$

$$T_L = 134/24$$

These selection probabilities were used in order to give each child in one of the S, M, or L subuniverses an equal probability of selection, since approximately one class of children per center was tested by PRC.

Sampling of centers was systematic with a random start (see Appendix A). The lists were ordered by region as well as program length in the hope that variance could be reduced. This was suggested by the Census Bureau's Mr. Robert Hanson, who was most generous with suggestions for the drawing of the PRC sample as well as details of the Census Bureau draw.

Those interested in the geographic distribution of full-year Head Start centers would be well advised to look at the quite complete description of the PRC universe given by the second Census Bureau sample listed in Exhibits A-1, A-2, and A-3 of Appendix A. The geographic distribution of centers in this universe is different from the textbook kind.

4. Replacement of Centers

A number of centers in the original draw had to be replaced, principally because they ended their program so much earlier than the early estimated date that they could not be included even with the information obtained by phone. (A great deal of rescheduling was successful because of the phoning of every center.) Some centers were also replaced because the children had been PPVT-tested two or more times already.

The first two centers were replaced ad hoc by centers under the same grant; however, since it soon became clear that a more easily explainable replacement scheme was needed, one was worked out. Each center in the second Census Bureau sample was given a replacement week number, which can differ by one from the ending week number since replacement weeks end on Thursday. Centers were then replaced, if necessary, by a second Census Bureau sample center in the same OEO region and with a replacement week number as large or larger than the replacement week number of the center to be replaced. Within these possibilities, probability of choice was as described in the previous section with a slight adjustment to overcome the bias because of the further conditioning.

5. Mechanics of Drawing Children

As explained in subsection III. C. 3, the selection of about 12 children per center resulted in each child in an S, M, or L subuniverse entry in the sample with the same probability.

In addition to avoiding bias, another PRC aim in the drawing of children was to include as many children as possible from the Census Bureau sample of children. Since the Census Bureau chose every fifth child in every sample center, PRC usually chose more, but sometimes fewer, children from a center. A detailed set of instructions was prepared and issued to the testers for the drawing of a systematic sample augmenting, diminishing, or, if necessary, ignoring the Census Bureau sample with the random start appropriate to a number of different possible lists or orderings of children. It was usually unnecessary for the testers to use these instructions. However, even when they were not used, they did serve the purpose of impressing on the testers that the choice was not theirs and that the testers should phone, if necessary, for instructions specific to the list at hand. For most centers it was possible (since the Census Bureau random start had been obtained by phone along with class number and size information) to provide the testers in advance with a list of 15 child identification numbers which were Census children, in addition to random choices from within classes. An additional five child ID's were provided as spares; in some cases, the testers had to phone for more. Testers had to explain all failures to test before being allowed to use spares.

D. Primary Data Collection

1. Selection and Training of Testers

Candidates for Head Start testers were drawn from several sources. They included Washington area universities, Boston area universities, the Personnel Placement Office of the Peace Corps, Bryn Mawr College, the Professional Placement Office of the U.S. Employment Service, and various Planning Research Corporation contacts. Approximately 200 candidates were interviewed. Emphasis was placed on selecting those individuals who had the following qualifications:

- A college degree in a field related to education, sociology, psychology, or guidance.
- Experience with pre-schoolers or disadvantaged children.
- Language fluency in Spanish as well as English.
- A flexible summer schedule.

Each individual hired was required to provide a recommendation from one of his college instructors.

Twenty-five testers were trained in two training sessions (12-15 May and 8-11 June 1966) held at the Center of Adult Education at the

University of Maryland. Both training programs lasted for 3-1/2 days and covered the following:

- Orientation to the project and to test administration.
- Tester-child relationships with emphasis on the culturally deprived child.
- Adult-adult relationships.
- Role-playing in possible test situations.
- Procedures in data collection and scoring.
- Practice in testing young children.
- Practice in interviewing.

Testers worked from 1 to 7 weeks in the field but not continuously. Some testers were able to devote a major portion of their summer to the testing program and thus provided a corps of trained testers with considerable experience in testing Head Start children. Quality and uniformity in the testing program were controlled by frequent field supervision, as explained below. (See Appendix C for a more detailed description of the selection, training, and supervision of testers.)

2. Individual Tests

The three individual tests were always given in the following order: PSI, PPVT, and DAP. The most quiet and private places available were used for the individual tests. The tester began the testing by taking time to put the child at ease and establish rapport. The child received all three tests in one sitting, and the tester allowed time for "breaks" and relaxation during testing and between tests.

Of course some children were unable to be tested. If a child would not respond to the tester, despite efforts to establish rapport and gain the child's cooperation, a notation was made on his record that he was "untestable" and another child was selected for testing. Testers generally devoted from 20 to 60 minutes to a child before deciding that he was untestable. (See page 23 and Exhibits 4 and 5 for a description of untested and untestable children.)

3. Rating Scales

After the tester had tested a child, he arranged for an interview with the teacher (or teacher-aide) about the child. During the interview, which was carefully scheduled to elicit information required by the scale, the tester completed the Vineland Social Maturity Scale. The tester also gave the teacher a blank BI for each child tested and encouraged the teacher to make the ratings that afternoon or evening.

The tester was responsible for the collection of the 5-piece set of data on each child (PSI, PPVT, DAP, BI, and VSMS), as well as for the identification of the children to be tested. The latter was based on a sampling procedure provided the tester. In no case known did testers select or test children on any basis other than the sampling procedures given them, as described above.

4. Testing Supervision

In order to ensure a high quality of testing, each tester was observed at least once and in many cases several times during his field work. Two PRC staff members (Naomi H. Henderson and Ann O'Keefe) with extensive experience in testing and field testing served as field supervisors. They observed the testers during actual test situations, and they sat in on Vineland interviews with teachers. No comments or interruptions were made during the actual testing or interviewing; however, upon conclusion of a session, the supervisor conferred with the tester privately and emphasized those facets of test administration which would increase the uniformity and quality of the tester's testing. Observations of tester-child situations were arranged so that the supervisor was out of the line of vision of the child, and thus detracted as little as possible from a normal test situation.

5. Spanish-Speaking Children

There were enough Spanish-speaking children in the sample to warrant special testing procedures. The presence of Spanish-speaking children in a CDC was ascertained, and testers fluent in

Spanish were sent to those locations. When a tester encountered a Spanish-speaking child, he administered the tests in Spanish[1] and an appropriate notation was made on the child's test data. (See Exhibit 31 for a description of the sample's composition in terms of children who were tested in Spanish because their primary language was Spanish.)

E. Supplementary Data Collection

Since the sample of CDC's in which testing of children was conducted fell entirely within the national sample of CDC's defined and identified by the Bureau of the Census for its data collection, it was feasible to use information about staff members, programs, and families obtained in the Census Bureau's study. In particular, PRC depended upon the Census Bureau to provide data from the following:

- Staff Member Information Form (CAP-HS Form 43)
- Paid and Voluntary Workers' Evaluation Form (CAP-HS Form 38)
- Family Information Form (CAP-HS Form 46)

The Family Information Form was mailed by the Census Bureau to the parents of children in its sample. Additional mailings were made by the Census Bureau to the parents of children in PRC's sample who were not included in the Census Bureau sample.

---

[1] See Appendix D for the Spanish translations used.

IV. RESULTS

This section presents the results of the study with little or no discussion or interpretation. The data and analyses to be presented are intended to serve two purposes: (1) to support the fulfillment of the objectives of the study, and (2) to contribute to the pool of descriptive data about children, parents, and workers participating in Project Head Start, and about the instruments used to access them.

A. Composition of Resultant Samples

1. Children

As described in subsection III. C, samples of CDC's were drawn to represent programs of three different levels of program length. These three levels have been designated Short-term, Medium-term, and Long-term, and hereafter shall be referred to for convenience as S, M, and L. The initial designation of S, M, or L CDC's was based on records obtained from the Census Bureau, which listed each CDC in its sample as of 1 May 1966. The listing included starting and ending dates of each CDC.

As stated earlier, however, while length of program operation is a fairly good measure of the length of time that children have been in the program, the two variables are not perfectly correlated. Furthermore, it was found that individual programs in some cases had to modify or shift their operating schedules. These two factors have led PRC to (1) make adjustments in its classification of centers by S, M, or L, and (2) distinguish between two types of children in resultant samples.

a. Child Development Center Classification

CDC's were classified S, M, or L on the basis of the number of weeks the program had been in operation from opening day until the end of the week of testing. Centers classified S were those of 15 weeks' length or less. M type centers were those of 17 to 23 weeks' length; L type centers were 25 weeks or more in operating length as defined. Most CDC's were tested within 2 weeks of their closing time. In no case was a CDC tested more than 3 weeks prior to its closing date. Holidays (except for a 2-week period around Christmas) were not taken into account in calculation of the length of the program or the time in the program, nor was the daily length of the program considered.

A total of 72 CDC's was tested. The location and classification of these by S, M, and L are shown in Exhibit 1. It should be noted that one CDC fell between the intervals for S and M centers in program duration. The principal adjustments in classification of CDC's came with centers originally classified as L centers on the basis of the Census Bureau listing cited earlier. Three of the original 24 L centers were reclassified as M centers. The M center that fell between the S and M intervals was dropped from consideration for analyses based on an S, M, or L classification. Two S centers were dropped from most statistical analyses, since all children in the center samples were tested in Spanish. Thus, for analyses of results by S, M, or L, there are a total of 69 CDC's, while for descriptive data on eligible children, the number of CDC's used is 71.

It should be noted that, throughout the report, centers are referred to by 2-digit numbers, ranging from 01 to 72. These numbers are consistent within the report, in that, for example, center 01 always refers to one particular center. However, they correspond in no systematic way to the list of centers in Exhibit 1. The new 2-digit center identification numbers were assigned to prevent particular centers from being singled out, and to help maintain anonymity of centers discussed in this report.

b. Eligible Versus Non-Eligible Children

In all, 964 children in the 72 CDC's were tested with the five tests and rating scales. These were, of course, children who (1) were identified by PRC's sampling procedure, (2) were available for testing during the week in which their CDC was tested, and (3)

EXHIBIT 1 - LOCATIONS OF SHORT-, MEDIUM-, AND LONG-TERM CENTERS

Short (15 Weeks or Less)	Medium (17 to 23 Weeks)	Long (25 Weeks or More)
Los Angeles, California	Cedar Creek, Missouri	Philadelphia, Pennsylvania
Los Angeles, California	Abilene, Texas	Chicago, Illinois
LaPuente, California	St. James, Missouri	Chicago, Illinois
Compton, California	Wilmington, Illinois	Chicago, Illinois
Pomona, California	Buffalo, New York	Chicago, Illinois
Mission, Texas	Cincinnati, Ohio	Chicago, Illinois
Talihina, Oklahoma	Keokee, Virginia	Chicago, Illinois
El Paso, Texas	Denver, Colorado	Mamaroneck, New York
Kansas City, Missouri	Modesto, California	Chicago, Illinois
Hampton, Virginia	Commerce City, Colorado	Chicago, Illinois
El Paso, Texas(1)	Eastman, Georgia	New Orleans, Louisiana
Paterson, New Jersey	Lincoln, Nebraska	New Orleans, Louisiana
El Paso, Texas(1)	Kirksville, Missouri	Lackawanna, New York
Montclair, New Jersey	Lincoln, Nebraska	Philadelphia, Pennsylvania
Waelder, Texas	Chicago, Illinois	Philadelphia, Pennsylvania
Silver City, New Mexico	Omaha, Nebraska	Philadelphia, Pennsylvania
Los Angeles, California	Riverbank, California	Philadelphia, Pennsylvania
Los Angeles, California	Philadelphia, Pennsylvania	Philadelphia, Pennsylvania
Venice, California	Owensboro, Kentucky	Philadelphia, Pennsylvania
Compton, California	Philadelphia, Pennsylvania	Philadelphia, Pennsylvania
Moorestown, New Jersey	Philadelphia, Pennsylvania	Philadelphia, Pennsylvania
Los Angeles, California	Avondale, Colorado	
Compton, California	Hamilton, Montana	
Los Angeles, California	Pittsfield, Massachusetts	
	Pittsfield, Massachusetts	
	Philadelphia, Pennsylvania	
N = 24	N = 26(2)	N = 21

Notes: (1) Not used in main analyses of variance and regression.

(2) One M center (Ridgefield, Washington) is not included due to its length of program (16 weeks). In some analyses, however, it is treated as an M center.

were found to be testable. However, there was substantial variation in the number of weeks that children in given programs had actually been in attendance. Since PRC was interested in the effect of the length of the programs, it was found necessary to classify the children into two groups: eligible and non-eligible.

The term "eligible" is used to designate that the child was in a CDC of the jth type (S, M, or L) and was in the center's program for the number of weeks falling within the jth type duration interval. Occasional illnesses or absences and holidays were not considered in calculations of eligibility. The term "non-eligible" or "dropped" is used to designate that the child tested had been in the program a fewer number of weeks (or in a few instances, a greater number of weeks) than those within the defined interval by which his center was classified. In effect, then, a child was given the S, M, or L label of his center for the purpose of most of PRC's analyses, but he was included in many of these analyses only if he was eligible. That is, in an analysis of effects by S, M, and L, a child from an M center who had only been in the program for a short time would not be included with S children; he would be dropped from the tabulation or analysis. The total samples, whether for a center or for S, M, or L type centers combined, were the sum of eligible and non-eligible children. For the total combined S, M, and L sets of centers there were 831 eligible children, of whom 67 were tested in Spanish.

c. Distribution of Children by Sex, Race, and Age

Exhibit 2 shows the composition by sex, race, and age of the S, M, and L samples, based on all children tested (i.e., eligible plus non-eligible). The percentages are percentages within each major sample. The number of children in each sex, race, and age category is given only for the finest breakdown to avoid cluttering the tables.

Exhibit 3 shows the composition by sex, race, and age of the S, M, and L samples, based on all eligible children tested. Generally

EXHIBIT 2 - SAMPLE COMPOSITIONS IN NUMBERS AND PERCENTAGES OF ALL CHILDREN TESTED, BY SEX, RACE, AND AGE(1)

		Male								Female							
Program Duration Category	Factor Classification	White				Non-White				White				Non-White			
	Ages	3/0-3/11	4/0-4/11	5/0-5/11	6/0-6/11	3/0-3/11	4/0-4/11	5/0-5/11	6/0-6/11	3/0-3/11	4/0-4/11	5/0-5/11	6/0-6/11	3/0-3/11	4/0-4/11	5/0-5/11	6/0-6/11
Short (N = 324)	S x R x A(2)	0.93 (3)	2.47 (8)	6.48 (21)	7.72(3) (25)	2.78 (9)	16.05 (52)	10.49 (34)	1.54 (5)	1.23 (4)	5.56 (18)	7.41 (24)	7.72 (25)	1.85 (6)	15.74 (51)	8.02 (26)	4.01 (13)
Medium (N = 353)	S x R x A	0.28 (1)	12.46 (44)	17.56 (62)	9.63(3) (34)	2.27(4) (8)	8.22 (29)	3.69 (13)	.57 (2)	1.13 (4)	10.48 (37)	14.16 (50)	6.80 (24)	2.27 (8)	5.10 (18)	4.53 (16)	.85 (3)
Long (N = 287) (N_T = 964)	S x R x A	0.35 (1)	4.20 (12)	3.83 (11)	- (0)	6.27 (18)	23.34 (67)	12.20 (35)	.35 (1)	.70 (2)	1.74 (5)	2.79 (8)	- (0)	3.14 (9)	23.69 (68)	17.07 (49)	.35 (1)
Short	S x R	18.22				30.86				21.61				29.31			
Medium	S x R	39.93				14.75				32.57				12.75			
Long	S x R	8.38				41.81				5.23				44.25			
Short	R x A	2.16	8.34	13.89	15.44	4.63	31.48	18.51	5.55								
Medium	R x A	1.41	22.94	31.72	16.43	4.54	13.32	8.22	1.42								
Long	R x A	1.05	5.94	6.62	-	9.06	47.03	29.27	.70								
Short	A	6.79	39.82	32.40	20.99												
Medium	A	5.95	36.26	39.94	17.85												
Long	A	10.11	52.69	35.89	.70												
Short	R	39.83				60.17											
Medium	R	72.50				27.50											
Long	R	13.61				86.06											
Short	S	49.08								50.92							
Medium	S	54.68								45.32							
Long	S	50.19								49.48							

Notes: (1) Details are percentages of children in a given classification for the indicated program duration category. Frequencies are shown in parentheses for the Sex x Race x Age classification. Percentages may not total 100 due to rounding. Samples include children tested in Spanish.

(2) S x R x A indicates Sex x Race x Age.

(3) One child included here was over 6/11.

(4) One child included here was under 3/0.

EXHIBIT 3 - SAMPLE COMPOSITIONS IN NUMBERS AND PERCENTAGES OF "ELIGIBLE" CHILDREN ONLY, BY SEX, RACE, AND AGE(1)

Program Duration Category	Factor Classification	Male								Female							
		White				Non-White				White				Non-White			
		Age 3/0-3/11	4/0-4/11	5/0-5/11	6/0-6/11	3/0-3/11	4/0-4/11	5/0-5/11	6/0-6/11	3/0-3/11	4/0-4/11	5/0-5/11	6/0-6/11	3/0-3/11	4/0-4/11	5/0-5/11	6/0-6/11
Short (N = 324)	S x R x A(2)	.93 (3)	2.47 (8)	6.48 (21)	7.72(3) (25)	2.78 (9)	16.05 (52)	10.49 (34)	1.54 (5)	1.23 (4)	5.56 (18)	7.41 (24)	7.72 (25)	1.85 (6)	15.74 (51)	8.02 (26)	4.01 (13)
Medium (N = 295)	S x R x A	.34 (1)	12.54 (37)	16.61 (49)	10.85(3) (32)	2.37(4) (7)	6.78 (20)	3.05 (9)	.68 (2)	1.36 (4)	10.51 (31)	14.58 (43)	8.14 (24)	2.37 (7)	5.08 (15)	3.73 (11)	1.02 (3)
Long (N = 212) (N_T = 831)	S x R x A	0.00	4.72 (10)	3.77 (8)	- (0)	5.19 (11)	25.47 (54)	13.68 (29)	.47 (1)	.94 (2)	2.36 (5)	2.83 (6)	- (0)	3.30 (7)	19.34 (41)	17.45 (37)	.47 (1)
Short	S x R	18.22				30.86				21.61				29.31			
Medium	S x R	40.34				12.88				34.92				11.86			
Long	S x R	8.49				44.34				6.13				40.56			
Short	R x A	2.16	8.34	13.89	15.44	4.63	31.48	18.51	5.55								
Medium	R x A	1.70	23.05	31.19	19.32	4.74	11.86	6.78	1.36								
Long	R x A	.94	7.08	6.60	-	8.02	44.81	31.13	.94								
Short	A	6.79	39.82	32.40	20.99												
Medium	A	6.44	34.91	37.97	20.68												
Long	A	8.96	51.89	37.73	.94												
Short	R	39.83				60.17											
Medium	R	75.26				24.74											
Long	R	14.62				84.90											
Short	S	49.08								50.92							
Medium	S	53.22								46.78							
Long	S	52.83								46.69							

Notes: (1) Details are percentages of children in a given classification for the indicated program duration category. Frequencies are shown in parentheses for the S x R x A classification. Percentages may not total 100 due to rounding. Sample includes children tested in Spanish.

(2) S x R x A indicates Sex x Race x Age.

(3) One child included here was over 6/11.

(4) One child included here was under 3/0.

in this report the classification "non-white" refers to Negro children. However, included in this classification are also two Oriental children and nine Indian children.

d. Distribution of Children by Language Used in Testing

Exhibit 4 shows the composition by sex, age, and primary language (English or Spanish) of the S, M, and L samples, based on all eligible children tested.[1]

e. Untestable and Untested Children

As mentioned earlier (see page 14), some children in the original sample were unable to be tested. In some cases, the children were simply unavailable due to absence. In other cases, the child was untestable because he would not respond to the tester in the test situation. Testers generally spent from 20 to 60 minutes attempting to gain a child's cooperation before deciding to consider the child untestable.

Exhibit 5 presents a description in terms of sex, race, and age (where such information was available) of the number of children who were untestable (UT) or absent (A). Absent children included children who had never attended the program but had been registered, children who were ill, children who had been withdrawn from the program for a variety of reasons, and children whose families had moved. Untestable children included children who (1) refused to go with the tester to the testing room, (2) were extremely reticent, (3) were tearful and uncommunicative, (4) spoke unintelligibly, and (5) were unmanageable and hyperactive.

Of the children who were selected for inclusion in the test sample, 60 were eventually deemed untestable and 257 were found to be absent or unavailable during the week of testing at their center.

---

[1] When a child's primary language was Spanish, he was almost always tested in Spanish by a tester fluent in the Spanish language. (See Appendix D for Spanish version of tests.) However, if a child's primary language was Spanish, but he was tested in English because in the tester's judgment he was not unduly restricted by the language, that child was not classified as "Spanish-speaking," "tested in Spanish," or "Spanish."

EXHIBIT 4 - SUMMARY OF ELIGIBLE CHILDREN, BY PROGRAM DURATION, SEX, AGE, AND LANGUAGE USED IN TESTING

Sex	Language	Short					Medium					Long					Total
		Age					Age					Age					
		3	4	5	6	Sub-Total	3	4	5	6	Sub-Total	3	4	5	6	Sub-Total	
Male	English(1)	12	60	50	7	129	8	55	58	33	154	11	61	37	1	110	-
	Spanish(2)	0	0	5	23	28	0	2	0	1	3	0	3	0	0	3	-
	Total(3)	12	60	55	30	157	8	57	58	34	157	11	64	37	1	113	429
Female	English	10	66	47	18	141	11	45	53	25	134	9	45	41	1	96	-
	Spanish	0	3	3	20	26	0	1	1	2	4	0	1	2	0	3	-
	Total	10	69	50	38	167	11	46	54	27	138	9	46	43	1	99	402
Male and Female	English	22	126	97	25	270	19	100	111	58	288	20	106	78	2	206	-
	Spanish	0	3	8	43	54	0	3	1	3	7	0	4	2	0	6	-
	Total	22	129	105	68	324	19	103	112	61	295	20	110	80	2	212	831

Notes: (1) English is primary language (combines white and non white).
(2) Spanish is primary language (white only).
(3) Total (English and Spanish).

EXHIBIT 5 - NUMBER OF UNTESTABLE AND UNTESTED (ABSENT) CHILDREN, BY SEX, RACE, AND AGE

Age	Reason for No Test	Male			Female			Unknown(2)	Total Untestable	Total Absent	Total
		White	Non-White	Unknown(1)	White	Non-White	Unknown(1)				
3	UT(3)	4	6	1	2	2	-	-	15	-	-
	A(4)	5	4	1	4	5	-	-	-	19	34
4	UT	6	9	2	5	4	-	-	26	-	-
	A	12	24	6	10	34	9	-	-	95	121
5	UT	4	1	1	3	2	-	-	11	-	-
	A	10	29	3	11	18	6	1	-	78	89
6	UT	1	-	3	2	-	1	-	7	-	-
	A	7	2	3	6	1	-	3	-	22	29
Unknown(5)	UT	-	-	-	1	-	-	-	1	-	-
	A	1	9	3	2	2	8	18	-	43	44
	Total UT	15	16	7	13	8	1	0	60	-	-
	Total A	35	68	16	33	60	23	22	-	257	-
	Total	50	84	23	46	68	24	22	-	-	317

Notes: (1) Children for whom race was not indicated.

(2) Children for whom neither race nor sex was indicated.

(3) UT indicates untestable.

(4) A indicates absent or unavailable.

(5) Children for whom age was not indicated.

Exhibit 6 shows the percentages of untestable children by sex, race, and age. In this exhibit, only untestable children for whom race, sex, and age were indicated (N = 51) are included. (If all 60 untestable children are included, the total percent of children considered untestable rises from the 5.02 percent indicated in Exhibit 6 to about 5.9 percent.[1]) Exhibits 5 and 6 show that 3- and 4-year-old children were more likely to be considered untestable than 5- and 6-year-old children. It also shows that boys tended to be more untestable than girls, and whites more than non-whites. In effect, the children with the highest untestable rates were young white boys.

2. Staff Members - Distributions by Characteristics

In this subsection, the total staff member population for S, M, and L centers will be described in terms of their position, selected characteristics, and relevant experience. All data are from the Staff Member Information Forms (CAP-HS Form 43) collected from the selected Child Development Centers) by the Bureau of the Census. Head Start workers were classified according to the following four positions:

- Administrative workers
- Teachers
- Other professionals (psychologists, physicians, dentists, and social workers)
- Non-professionals (paid teachers' aides and volunteers)

Exhibit 7 presents percentages of staff members by position for S, M, and L centers. In S centers, non-professional workers made up about 67 percent of the total staff, with teachers accounting for the next 21 percent. This same distribution pattern prevailed in M and L centers, with percentages reported for each staff position being proportionately the same. In all three levels, the administrative workers comprised the smallest percent of the total staff population.

---

[1] In all, 1,024 children were approached for testing, 964 were tested, and 60 were considered untestable.

EXHIBIT 6 - PERCENTAGE OF UNTESTABLE CHILDREN BY SEX, RACE, AND AGE

Factor Classification	Male						Female					
	White			Non-White			White			Non-White		
	No. Tested	No. UT(1)	% UT(1)	No. Tested	No. UT(1)	% UT(1)	No. Tested	No. UT(1)	% UT(1)	No. Tested	No. UT(1)	% UT(1)
Age 3	5	4	44.44	35	6	14.63	10	2	16.66	23	2	8.00
Age 4	64	6	8.57	148	9	5.73	60	5	7.69	137	4	2.83
Age 5	94	4	4.08	82	1	1.20	82	3	3.52	91	2	2.15
Age 6	59	1	1.66	8	0	0.00	49	2	3.92	17	0	0.00
Sex/Race	222	15	6.32	273	16	5.53	201	12	5.63	268	8	2.89
Race	423	27	6.00	541	24	4.24						
Sex	495	31	5.89				469	20	4.08			
Total	964	51	5.02									

Note: (1) UT indicates untestable. The "percent UT" is based on the total number of children attempted of the type indicated; that is, UT/number tested + UT.

EXHIBIT 7 - PERCENTAGE OF STAFF MEMBERS BY STAFF POSITION, FOR SHORT-, MEDIUM-, AND LONG-TERM CENTERS(1)

Length of Program	Type of Position	Total	Not Reported
Short	Administrative Workers	4.3	-
	Teachers	20.8	-
	Other Professionals	8.0	-
	Non-Professionals	66.7	-
	Not Answered	0.2	-
	Total	100.0	-
Medium	Administrative Workers	7.4	-
	Teachers	21.5	-
	Other Professionals	12.9	0.6
	Non-Professionals	58.3	1.3
	Not Answered	0.0	-
	Total	100.0	1.9
Long	Administrative Workers	5.6	-
	Teachers	30.2	-
	Other Professionals	8.1	-
	Non-Professionals	55.1	-
	Not Answered	1.0	-
	Total	100.0	-

Reported information on age of staff members is shown in Exhibit 8. Percent figures indicate age by staff position for S, M, and L centers. The modal age of administrative staff in all center types was 45 years or over. In both S and L centers, teachers most frequently ranged from 22 to 30 years in age and accounted for approximately 45 percent of all teachers employed. In M centers, however, the majority of teachers were between 31 and 45 years in age. With respect to other professionals, about 14 percent in M centers were under age 22, while S and L centers listed no "other professionals" at this age level. The age range most frequently reported for non-professionals was 31 to 45 years, but (as would be expected) a significant percentage in all center types was under age 22.

Cultural and ethnic backgrounds of staff workers were also examined. Eight ethnic and cultural backgrounds were considered and Exhibit 9 lists resulting percentage distributions by staff position and center duration. About 17 percent of the S staff was Mexican/American; this figure was influenced by the fact that a great majority of all Spanish-speaking children were enrolled in S centers. In both S and M centers, percentage distributions for total staff population were highest for whites, and Negroes had the highest percentage in L centers. However, it is clear that the relative proportion of Negroes and whites employed in each position varied considerably among the three categories of program duration. Negroes, for example, occupied the majority of administrative positions, except in M centers, where whites comprised 85 percent of the administrative staff. White teachers, on the other hand, were most frequently reported for all three categories of center duration, but sample percentages reported vary greatly. The majority of non-professional staff members reported in S and L centers was Negro.

Exhibit 10 presents information on paid and volunteer staff members. Paid workers accounted for at least 76 percent of the total staff in each program level. Except in L centers, volunteers most frequently reported were the "other professionals." Teachers and administrators

EXHIBIT 8 - AGE DISTRIBUTIONS OF STAFF MEMBERS BY STAFF POSITION, FOR SHORT-, MEDIUM-, AND LONG-TERM CENTERS(1)

Program Duration	Type of Position	Total(2)	Age				
			Under 22	22-30	31-45	Over 45	Not Reported
Short	Administrative Workers	100.0	-	-	47.0	53.0	-
	Teachers	100.0	1.7	44.5	23.6	30.5	-
	Other Professionals	100.0	-	28.8	36.4	23.4	10.9
	Non-Professionals	100.0	12.5	13.9	44.8	27.2	1.6
	Total	100.0	8.8	20.8	39.7	28.7	1.9
Medium	Administrative Workers	100.0	-	8.9	19.9	71.2	-
	Teachers	100.0	2.9	32.0	41.0	24.1	-
	Other Professionals	100.0	13.7	20.7	33.9	27.1	4.9
	Non-Professionals	100.0	15.2	27.3	25.7	29.6	2.2
	Total	100.0	11.3	26.1	29.6	31.1	1.9
Long	Administrative Workers	100.0	-	6.9	40.5	45.7	6.9
	Teachers	100.0	1.6	45.6	37.2	7.4	8.0
	Other Professionals	100.0	-	13.2	38.9	35.9	12.0
	Non-Professionals	100.0	22.3	23.6	27.6	23.4	3.2
	Total	100.0	12.8	28.2	32.8	20.6	5.6

Notes: (1) Details are percentages of stub totals.

(2) Percentages may not add to 100 due to rounding.

EXHIBIT 9 - CULTURAL AND ETHNIC CHARACTERISTICS OF STAFF MEMBERS BY STAFF POSITION, FOR SHORT-, MEDIUM-, AND LONG-TERM CENTERS(1)

Program Duration	Type of Position	Total	White	Negro	Ori-ental	Amer. Ind.	Mex./Amer.	Puerto Rican	Es-kimo	Other	Not Re-ported
Short	Administrative Workers	100.0	46.0	51.0	-	-	4.0	-	-	-	-
	Teachers	100.0	53.4	28.8	-	-	17.7	-	-	-	-
	Other Professionals	100.0	87.0	10.9	-	-	2.2	-	-	-	-
	Non-Professionals	100.0	34.1	40.4	-	4.7	19.6	-	-	1.1	-
	Total	100.0	42.8	36.2	-	3.1	17.2	-	-	0.7	-
Medium	Administrative Workers	100.0	85.2	12.7	-	-	2.1	-	-	-	-
	Teachers	100.0	71.8	17.5	-	5.8	-	-	-	4.8	-
	Other Professionals	100.0	68.3	24.4	2.0	2.7	2.7	-	-	-	-
	Non-Professionals	100.0	55.4	30.4	2.2	4.9	4.4	-	-	2.7	-
	Total	100.0	62.8	25.6	1.5	4.5	3.1	-	-	2.6	-
Long	Administrative Workers	100.0	39.7	60.3	-	-	-	-	-	-	-
	Teachers	100.0	49.8	40.1	-	-	-	-	-	10.1	-
	Other Professionals	100.0	47.3	47.3	-	4.8	-	-	-	-	-
	Non-Professionals	100.0	18.4	60.7	-	7.7	0.9	-	-	12.2	-
	Total	100.0	31.3	53.8	-	4.7	0.5	-	-	9.8	-

Note: (1) Details are percentages of stub totals.

EXHIBIT 10 - PERCENTAGE OF PAID VERSUS VOLUNTEER STAFF MEMBERS BY STAFF POSITION, FOR SHORT-, MEDIUM-, AND LONG-TERM CENTERS

Program Duration	Type of Position	Total(1)	Paid	Volunteer
Short	Administrative Workers	100.0	87.0	13.0
	Teachers	100.0	88.9	8.4
	Other Professionals	100.0	40.2	59.8
	Non-Professionals	100.0	75.4	24.6
	Total	100.0	76.0	23.5
Medium	Administrative Workers	100.0	66.1	33.9
	Teachers	100.0	91.2	8.8
	Other Professionals	100.0	45.9	54.1
	Non-Professionals	100.0	79.1	20.5
	Total	100.0	76.5	23.3
Long	Administrative Workers	100.0	100.0	-
	Teachers	100.0	98.7	1.3
	Other Professionals	100.0	100.0	-
	Non-Professionals	100.0	79.6	20.4
	Total	100.0	88.4	11.6

Note: (1) Percentages may not add to 100 due to rounding or not being answered.

held the highest percentage of paid positions, except in M centers, where teachers and non-professionals comprised the greatest percentage of the paid staff.

Data on staff experience with pre-school children are presented in Exhibit 11. Again, figures listed are percentages by staff position for S, M, and L centers. The information reported indicates that the greatest percentage of the total staff population in both S and M centers had no previous experience with pre-school children. For administrative staff, a minimum of 5 years' experience was most frequently reported in all levels. Teachers most frequently had from 1 to 3 years' experience while the modal percentage of non-professionals in two of the three levels had not worked with pre-school children before. The lower age level of the non-professionals is probably a prime factor contributing to that difference.

In Exhibit 12 a second analysis is made on the basis of staff experience with children from poverty conditions. Again, percents represent staff positions for S, M, and L centers. Using these criteria, it can be seen that over one-third of the total staff employed in each center type had no previous experience with children from conditions of poverty.

Overall percentage distributions by staff position within each level are much the same as those presented in Exhibit 11. Again the majority of administrators had over 5 years' experience and teachers most frequently reported from 1 to 3 years' experience related specifically to children from poverty backgrounds. The percentage of non-professionals reporting no experience with children from conditions of poverty included well over one-third of the total non-professionals employed.

3. Families - Distributions by Characteristics

Descriptive information about Head Start families was obtained by the Bureau of the Census from the Family Information Form (CAP-HS Form 46). Since many questionnaires were either

EXHIBIT 11 - NUMBER OF YEARS OF EXPERIENCE WITH PRE-SCHOOL CHILDREN BY STAFF POSITION, FOR SHORT-, MEDIUM-, AND LONG-TERM CENTERS

Program Duration	Type of Position	Total(2)	None	1-3 Years	4-5 Years	Over 5 Years	Not Answered
Short	Administrative Workers	100.0	-	8.0	4.0	88.0	-
	Teachers	100.0	21.9	45.9	7.3	25.1	-
	Other Professionals	100.0	17.9	27.7	-	54.3	-
	Non-Professionals	100.0	40.9	24.6	8.5	24.1	2.0
	Total	100.0	33.2	28.5	7.3	29.5	1.5
Medium	Administrative Workers	100.0	16.9	40.3	-	42.8	-
	Teachers	100.0	19.3	41.2	12.8	26.7	-
	Other Professionals	100.0	27.8	39.0	2.4	30.7	-
	Non-Professionals	100.0	48.5	30.0	4.7	14.8	2.0
	Total	100.0	37.2	34.3	5.8	21.5	1.2
Long	Administrative Workers	100.0	-	34.5	-	65.5	-
	Teachers	100.0	25.7	37.1	14.0	18.3	4.8
	Other Professionals	100.0	26.9	29.9	4.2	31.1	7.2
	Non-Professionals	100.0	32.0	37.6	4.8	22.8	2.9
	Total	100.0	27.6	36.8	7.2	24.8	3.7

Notes: (1) Details are percentages of stub totals.

(2) Percentages may not add to 100 due to rounding.

EXHIBIT 12 - NUMBER OF YEARS OF EXPERIENCE WITH CHILDREN FROM CONDITIONS OF POVERTY, BY STAFF POSITION, FOR SHORT-, MEDIUM-, AND LONG-TERM CENTERS(1)

Program Duration	Type of Position	Total(2)	None	1-3 Years	4-5 Years	Over 5 Years	Not Answered
Short	Administrative Workers	100.0	-	22.0	17.0	61.0	-
	Teachers	100.0	25.7	44.1	2.7	27.3	-
	Other Professionals	100.0	41.8	25.5	3.8	29.3	-
	Non-Professionals	100.0	47.1	21.6	9.1	19.7	2.5
	Total	100.0	40.1	26.6	7.7	23.8	1.8
Medium	Administrative Workers	100.0	5.5	18.2	4.2	71.6	-
	Teachers	100.0	16.8	58.2	5.8	19.1	-
	Other Professionals	100.0	44.1	26.3	10.0	19.5	-
	Non-Professionals	100.0	54.3	21.9	7.3	13.1	3.4
	Total	100.0	41.3	30.0	7.1	19.6	2.0
Long	Administrative Workers	100.0	6.0	13.8	17.2	62.9	-
	Teachers	100.0	29.7	54.7	2.9	11.7	1.0
	Other Professionals	100.0	26.9	21.0	16.8	35.3	-
	Non-Professionals	100.0	46.8	40.7	2.6	8.2	1.8
	Total	100.0	37.8	41.4	4.7	14.9	1.3

Notes: (1) Details are percentages of stub totals.

(2) Percentages may not add to 100 due to rounding.

incomplete or not returned, all resulting data are based on the returns of 656 families.[1] Exhibit 13 shows the percentage of families responding for (1) all sample children in the 71 S, M, and L centers, based on the assumption of a family per child in the sample.

Exhibit 14 summarizes the overall data relevant to both number of siblings and structural characteristics for families of children in S, M, and L centers. The average number of siblings in the S, M, and L center families was 3.76, 3.83, and 3.39, respectively. The information reported also indicates that for families of children in all three types of centers over 50 percent of the siblings were in the 6- to 15-year age range, while children under 6 years of age accounted for approximately 37 percent of the siblings.

Mean family size varies slightly among different subsamples. Exhibit 15 (page 39) presents the mean family size for three subgroups of children: eligible English-speaking children, eligible English-speaking children living with their mother and father, and children tested in Spanish. For the first group--English-speaking eligible children (regardless of whether or not a father was living at home)--the mean family size was 6.69 people. The second group consisted only of English-speaking children who lived with both father and mother; as might be expected, the mean family size was slightly higher, 7.71 people. For purposes of comparison, a separate mean family size was also calculated for a third small group--the 42 children who were tested in Spanish (regardless of whether or not they lived with both parents)--and was found to be 7.02 people.

In regard to family structural characteristics, Exhibit 14 shows that over 96 percent of the families in all three program types reported

---

[1] The Family Information Form unfortunately has a complicated-appearing format and uses somewhat sophisticated language. It is quite likely that some information reported on this form may therefore be inaccurate, since responding families may have at times misread or misinterpreted items or instructions. Furthermore, no reliability check on the reported data was made.

EXHIBIT 13 - PERCENTAGE OF FAMILY RESPONSE BASED ON SAMPLE CHILDREN

Program Type	Total Sample
Short	69.1
Medium	69.1
Long	65.5

EXHIBIT 14 - SIBLINGS AND FAMILY STRUCTURAL CHARACTERISTICS

Characteristic	Program Duration		
	Short	Medium	Long
Total number of families responding	224	244	188
Siblings			
Total number reported	843	934	638
Average number per family	3.76	3.83	3.39
Structure		Percent	
Siblings under 6 years old	37.72	37.58	36.99
Siblings from 6 to 15 years old	56.82	55.14	52.98
Siblings from 16 to 21 years old	5.34	6.64	9.09
Siblings over 21 years	0.12	0.64	0.94
Families with siblings living at home	91.50	97.50	93.10
Families with mother living with child	96.00	99.60	97.90
Families with father living with child	71.00	76.60	69.10
Families with other relatives living in home	14.30	6.10	14.90
Families receiving public welfare	29.00	31.00	25.00

that the mother was living at home with the child; the percentage of fathers living at home ranged from 69 to 76 percent among the three center types. These percentages, of course, take into account only families from a portion of the total sample. Exhibit 16 also shows the percentage of families reporting fathers living at home, but percentages in this exhibit are based on the actual numbers of total and eligible sample children in each program type. A comparison of these percentage figures with those given in Exhibit 14 shows that when the total sample is considered, the percentage of families with a father living at home decreases considerably.

The data presented in Exhibit 14 also include information on families receiving public welfare. Again, figures given are percentages and show that from 25 to 31 percent of the reporting families in all three center types were receiving public welfare aid.

The level of education completed by parents or guardians (given separately for fathers and mothers) of children in S, M, and L centers is presented in Exhibit 17. It is interesting that while 165 families did not report the father's level of education, only three families did not report this information for mothers. For both fathers and mothers, an educational level of at least 9 years was most frequently reported. Overall percentage distributions in both tables show that for the majority of all parents reported, the educational level ranged from 7 to 12 years of school. The highest average incidence of parents reporting no formal education was 3 percent for mothers of S children and 2 percent for fathers of M children. On the other hand, the highest average incidence of parents who had completed at least some college was 12 percent, a figure reported among both mothers and fathers of S children.

Family income data are presented in Exhibit 18, which lists percentage figures for families according to program type and level of income. The range of possible annual incomes is from less than $1,000 to $10,000 or more. As can be seen from the distribution, the majority of all families responding reported incomes ranging from

EXHIBIT 15 - MEAN FAMILY SIZE FOR THREE SUBGROUPS OF CHILDREN TESTED

Group Type	Total N	Mean Family Size
Eligible non-Spanish-speaking children	517	6.69
Eligible non-Spanish-speaking children with mother and father	387	7.11
Children tested in Spanish	42	7.02

EXHIBIT 16 - PERCENTAGE OF FAMILIES OF SAMPLE CHILDREN WITH FATHER LIVING AT HOME

Program Duration	Total Sample	Eligible Only
Short	49.1	49.1
Medium	53.0	63.4
Long	45.3	61.3

EXHIBIT 17 - LEVEL OF EDUCATION COMPLETED BY PARENTS OF CHILDREN IN SHORT-, MEDIUM-, AND LONG-TERM CENTERS

Grade Level of Education Completed	Father or Male Guardian						Mother or Female Guardian					
	Short		Medium		Long		Short		Medium		Long	
	N	%	N	%	N	%	N	%	N	%	N	%
No Formal Education	3	1.3	5	2.0	0	0.0	7	3.1	4	1.6	1	0.5
Elementary, 1-3	8	3.6	12	4.9	5	2.7	8	3.6	0	0.0	2	1.1
Elementary, 4-6	21	9.4	15	6.1	11	5.9	18	8.0	17	7.0	10	5.3
Elementary, 7-8	27	12.1	35	14.3	21	11.2	27	12.1	52	21.3	16	8.5
High School, 9-11	51	22.8	70	28.7	46	24.5	70	31.3	92	37.7	77	41.0
High School, 12	30	13.4	35	14.3	36	19.1	64	28.6	68	27.9	69	36.7
College (some or all)	27	12.1	17	7.0	16	8.5	27	12.1	11	4.5	13	6.9
Not Reported	57	25.4	55	22.5	53	28.2	3	1.3	0	0.0	0	0.0
Total(1)	224	100.1	244	99.8	188	100.1	224	100.1	244	99.9	188	100.0

Note: (1) Percentages may not equal 100 due to rounding.

EXHIBIT 18 - ANNUAL INCOME OF FAMILIES WITH CHILDREN IN SHORT-, MEDIUM-, AND LONG-TERM CENTERS(1)

Income Level	Short		Medium		Long	
	N	%	N	%	N	%
Less than $1,000	17	7.6	21	8.6	14	7.4
$1,000-$1,999	32	14.3	41	16.8	23	12.2
$2,000-$2,999	26	11.6	36	14.8	30	16.0
$3,000-$3,999	52	23.2	51	20.9	32	17.0
$4,000-$4,999	51	22.8	39	16.0	23	12.2
$5,000-$5,999	24	10.7	31	12.7	37	19.8
$6,000-$7,999	14	6.3	20	8.2	23	12.2
$8,000-$9,999	2	0.9	3	1.2	4	2.1
$10,000 or more	5	2.2	2	0.8	2	1.1
Not reported	1	0.4	0	0.0	0	0.0
Total	224	100.0	244	100.0	188	100.0

Note: (1) Data were obtained from Family Information Forms sent to families with children in the 1966 full-year Head Start program.

$2,000 to $5,000. The most frequently reported income was in the $3,000 to $4,000 range. Reported information also shows that, while few families listed incomes of $6,000 or more, roughly 8 percent of the families in the programs had incomes of less than $1,000.

B. Test Results

1. Distributions of Raw Scores and Unadjusted Means by Short-, Medium-, and Long-Term Centers

Exhibits 19A through 22C present raw score distributions on each of the four main tests for all 964 children tested in S, M, and L centers. Dropped or non-eligible children have been designated by an asterisk. Within a treatment classification, the distributions have been classified according to sex, age, and race.

Exhibits 19A, 19B, and 19C gives the distributions of raw scores for the Peabody Picture Vocabulary Test (PPVT), Form B. For this test the raw score is the ceiling point minus the number of items wrong. Exhibits 20A, 20B, and 20C give the distributions of raw scores of children for the revised Pre-School Inventory (PSI). Here the raw score is the total number of points obtained for the whole test, or the sum of the scores of the four subtests. The range of possible scores for the PSI total is from 0 to 90.

Exhibits 21A, 21B, and 21C present the distributions of raw s scores for the Behavior Inventory (BI). These are BI total scores, obtained by the addition of the total number of points for positive items to the total number of points for negative items. The range of possible scores is from 50 to 200.

Exhibits 22A, 22B, and 22C give the distributions of total raw scores for the Vineland Social Maturity Scale (VSMS). The total raw score on the VSMS is the basal score plus all points for which credit was received beyond the basal score.

The final six rows of Exhibits 19A through 22C indicate the numbers of children (N's), and mean and median raw scores for each sex-race-age classification, for both "total" children (N = 964, representing

EXHIBIT 19A - DISTRIBUTION OF PPVT RAW SCORES FOR CHILDREN IN SHORT-TERM CENTERS, BY SEX, RACE, AND AGE

	Male															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	28	27	25	27	26	28	31	22	14	15	13	22	35	35	44
		32	53	30	47	37	35	39		17	20	18	26	39	52	48
		33		35	48	41	38	40		22	25	22	27	41	54	
				40	50	41	39	42		22	26	24	27	44		
				41	51	41	41	43		24	26	29	27	48		
				43	52	43	42	44		33	27	30	28	49		
					55	45	42	45		42	27	30	31			
					56	49	45	50		42	28	30	33			
					56	57	47	55			30	31	33			
					61	57	48				32	31	34			
						59	48				34	31	36			
							49				37	33	37			
							52				37	33	37			
							54				39	34	38			
							60				40	35	39			
							62				40	35	39			
											42	37	39			
											45	39	40			
											46	39	41			
											46	39	43			
											51	40	43			
											54	41	46			
											58	48	48			
												52	50			
												53	53			
												53	55			
												54	58			
												54	68			
												57				
Total																
N	-	3	2	6	10	11	16	9	1	8	23	29	28	6	3	2
Mean	-	31.00	40.00	35.67	50.30	45.09	45.63	43.22	22.00	27.00	35.87	36.72	39.21	42.67	47.00	46.00
Median	-	32.00	40.00	37.50	51.50	43.00	46.50	43.00	22.00	23.00	37.00	35.00	38.50	42.50	52.00	46.00
Eligible																
N	-	3	2	6	10	11	16	9	1	8	23	29	28	6	3	2
Mean	-	31.00	40.00	35.67	50.30	45.09	45.63	43.22	22.00	27.00	35.87	36.72	39.21	42.67	47.00	46.00
Median	-	32.00	40.00	37.50	51.50	43.00	46.50	43.00	22.00	23.00	37.00	35.00	38.50	42.50	52.00	46.00

EXHIBIT 19A (Continued)

	Female															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	26	22	11	19	25	31	32	29	-	29	16	24	21	41	32	33
		33	27	20	31	32	32	38		29	18	24	25	51	40	36
		33	34	26	33	37	37	39		31	19	25	28	54	41	39
			45	29	34	40	37	42		31	24	26	29	57	43	40
				32	39	42	43	42		38	30	28	30		46	47
				35	42	43	43	42		40	30	28	31		50	
				36	43	50	44	42			32	30	31		52	
				38	44	50	44	43			33	31	33		54	
				38	45	52	46	45			34	31	34			
				41	45	53	47	46			34	31	35			
				44	46		52	58			34	31	37			
				47	46		53				36	32	39			
				48	52		55				38	33	40			
				52	56		58				40	38	42			
											41	38	43			
											53	39	43			
											56	39	43			
												39	44			
												41	44			
												41	48			
												41	48			
												42	56			
												42				
												42				
												44				
												44				
												44				
												46				
												46				
												47				
												47				
												50				
												51				
												53				
Total																
N	1	3	4	14	14	10	14	11	-	6	17	34	22	4	8	5
Mean	26.00	29.33	29.25	36.07	41.50	43.00	44.50	42.36	-	33.00	33.41	37.88	37.45	50.75	44.75	39.00
Median	26.00	33.00	30.50	37.00	43.50	42.50	44.00	42.00	-	31.00	34.00	39.00	38.00	52.50	44.50	39.00
Eligible																
N	1	3	4	14	14	10	14	11	-	6	17	34	22	4	8	5
Mean	26.00	29.33	29.25	36.07	41.50	43.00	44.50	42.36	-	33.00	33.41	37.88	37.45	50.75	44.75	39.00
Median	26.00	33.00	30.50	37.00	43.50	42.50	44.00	42.00	-	31.00	34.00	39.00	38.00	52.50	44.50	39.00

EXHIBIT 19B - DISTRIBUTION OF PPVT RAW SCORES FOR CHILDREN IN MEDIUM-TERM CENTERS, BY SEX, RACE, AND AGE

	Male															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	23	18 24*(1) 35 38 40* 46 48	26 28 28 31 31 32 36 37 38 39 41 41 41 42* 42 43* 43 43 43 46* 46 47 47 47 48* 48 50 51 52 52 53 53 54 56 58* 58 61	24 29 32 35 37 37 38* 43 43 43 43 44 45 45 45 46 47 48 49* 49 50 50 51 51 52 52 52 52 53* 54* 56* 56 57* 57* 57 58* 64 66	18 28 35 38 39 44 45 45* 45 46 47 49 49 51 51* 51 52* 52 53 54 54* 57 57* 58	42 45 45 45 49 50 50 51 51 52 53 54 54 54 55 55 55 56 57 58 58 59 59 61 63 64	40 42 45 46* 50 50 57 59*	24 42*	05 08 14 18 28 38	05 18 19 20 22 24* 27 28 29* 36* 44 46	20* 25* 26 27* 28 31 32 32 34* 39* 39 42 44* 45 47 48 48	24 26 34 34 36 37* 38 40* 42 44 45* 64*	48	43	44
Total																
N	-	1	7	37	38	24	26	8	2	6	12	17	12	1	1	1
Mean	-	23.00	35.57	44.11	47.63	46.58	53.65	48.63	33.00	18.50	26.50	39.88	38.67	48.00	43.00	44.00
Median	-	23.00	38.00	43.00	49.00	49.00	54.00	48.00	33.00	16.00	25.50	34.00	37.50	48.00	43.00	44.00
Eligible																
N	-	1	5	32	30	19	26	6	1	6	9	11	8	1	1	1
Mean	-	23.00	37.00	43.59	46.27	45.21	53.65	47.33	24.00	18.50	25.44	38.00	34.75	48.00	43.00	44.00
Median	-	23.00	38.00	43.00	46.50	47.00	54.00	47.50	24.00	16.00	22.00	39.00	35.00	48.00	43.00	44.00

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 17 weeks.

EXHIBIT 19B (Continued)

	Female															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	13 19 37 38	22 25 32 40 41 47 53	08 18*(1) 19 20 22 26 27 31 35 36 41 41* 43 44* 45 45 45 46 46* 46 46 47 47 47* 48 53 54* 55 55 56	09 28 28 30 31 33 34 36 36 40 41 41* 42 42 43 45 45* 48 48 48* 48* 49* 50 50* 52 52* 53 54 55 57	13 29 34 34 40 45 45 45 49 50 50 51 52 52 54 55 55 56 58 63	32 33 33 35 38 41 44 47 48 49 49 53 53 54 54 57 59	24 57 59 59 62 62 66	21*	13 17 18 20 24 25 28	21 25 26* 26 27	25 27 28 29 31 32* 35 36 41* 43 45 46 48	23 23* 24 28 33 34* 38* 41* 42 43* 44	28 46 49 54 55	61	27 46
Total																
N	-	4	7	30	30	20	17	7	1	7	5	13	11	5	1	2
Mean	-	26.75	37.14	39.73	42.27	46.50	45.82	55.57	21.00	20.71	25.00	35.85	33.90	46.40	61.00	36.50
Median	-	28.00	40.00	45.00	44.00	50.00	48.00	59.00	21.00	20.00	26.00	35.00	34.00	49.00	61.00	36.50
Eligible																
N	-	4	7	24	23	20	17	7	0	7	4	11	6	5	1	2
Mean	-	26.75	37.14	39.25	40.65	46.50	45.82	55.57	0.00	20.71	24.75	35.73	32.33	46.40	61.00	36.50
Median	-	28.00	40.00	45.00	42.00	50.00	48.00	59.00	0.00	20.00	25.50	35.00	30.50	49.00	61.00	36.50

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 17 weeks.

EXHIBIT 19C - DISTRIBUTION OF PPVT RAW SCORES FOR CHILDREN IN LONG-TERM CENTERS, BY SEX, RACE, AND AGE

	Male							
	White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	05*(1)	-	22	07 10* 18* 34 35 38 41 42 47 56 58	20 35* 43 46 51 52 54 55* 60	40 65*	-	-
Total								
N	1	-	1	11	9	2	-	-
Mean	5.00	-	22.00	35.09	46.22	52.50	-	-
Median	5.00	-	22.00	38.00	51.00	52.50	-	-
Eligible								
N	0	-	1	9	7	1	-	-
Mean	0.00	-	22.00	39.78	46.57	40.00	-	-
Median	0.00	-	22.00	41.00	51.00	40.00	-	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 19C (Continued)

	Male							
	Non-White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	30*[(1)] 35	11 17 18 19* 20* 20 21* 21 21* 26 37* 38 40 42 43 44*	19* 23 24* 24 27 30 34 35 36* 37 37 37 40 41 53	15 15 19 20* 23 23 24 25* 25 27 28 28 30 31 32* 32 32 33 34 35 36 36* 38 38 38 38 38* 39 39 40 40 41* 42 43 43 43 44* 44 45 46 47 47* 47 48* 48 49 49 51* 51 53 55 57* 59	22* 32 34 35 35 36* 36* 37 38 38 41* 41 41 42 42 42 43 43 43 43 44 45 45 45* 46 47 50 51 56 63 63	32 35 42* 45	52	-
Total								
N	2	16	15	53	31	4	1	-
Mean	32.50	27.38	33.13	37.79	42.55	38.50	52.00	-
Median	32.50	21.00	35.00	38.00	42.00	38.50	52.00	-
Eligible								
N	1	10	12	42	26	3	1	-
Mean	35.00	27.60	34.83	37.24	43.81	37.33	52.00	-
Median	35.00	23.50	36.00	38.00	43.00	35.00	52.00	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 19C (Continued)

	Female							
	White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	31 42	23 54	24 30 32	16 20 38 40*(1) 43* 49 51	44	-	-
Total								
N	-	2	2	3	7	1	-	-
Mean	-	36.50	38.50	28.67	36.71	44.00	-	-
Median	-	36.50	38.50	30.00	40.00	44.00	-	-
Eligible								
N	-	2	2	3	5	1	-	-
Mean	-	36.50	38.50	28.67	34.80	44.00	-	-
Median	-	36.50	38.50	30.00	38.00	44.00	-	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 19C (Continued)

	Female							
	Non-White							
	Age-Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	20	14	25	17*	25	44	47	-
	39	16	25*	19*	28	44		
	43	28*(1)	25	23	28	52		
		28	29*	24*	28	59		
		29	30	24	28			
		42*	30*	25	28			
			30*	27	28			
			31	27	30			
			33*	28	30			
			34	28*	30			
			34	31	30			
			36*	31	30*			
			37	33	31*			
			38*	33	31*			
			39*	34*	32			
			40	34	33			
			41	34*	34			
			42	34*	34			
			51*	36	35			
				36	35			
				36	37*			
				36	37*			
				37	37*			
				37*	37*			
				38	37			
				38*	38*			
				39	38			
				39*	39*			
				39*	40			
				39*	40			
				40	40			
				40	42*			
				40	42			
				42	42			
				43	44			
				43*	45			
				43*	48			
				43	50			
				43	52			
				43	52*			
				45	53			
				45*	53*			
				47*	54			
				47	55			
				48*	57			
				48				
				56				
				59				
Total								
N	3	6	19	48	45	4	1	
Mean	34.00	26.17	34.21	36.90	38.16	49.75	47.00	
Median	39.00	28.00	34.00	37.50	37.00	48.00	47.00	
Eligible								
N	3	4	10	31	33	4	1	
Mean	34.00	21.75	33.90	37.52	38.09	49.75	47.00	
Median	39.00	22.00	34.00	37.00	36.00	48.00	47.00	

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 20A - DISTRIBUTION OF PSI RAW SCORES FOR CHILDREN IN SHORT-TERM CENTERS, BY SEX, RACE, AND AGE

	Male															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	29	34	23	32	40	33	27	32	15	18	20	20	45	50	35
		31	43	38	51	41	38	45		18	21	22	26	46	68	72
		41		40	53	47	38	45		25	28	23	27	54	77	
				43	53	52	47	47		28	29	24	32	55		
				44	55	54	51	50		33	29	26	35	58		
				65	55	59	53	51		35	29	29	36	75		
					60	61	53	51		36	31	30	38			
					61	71	54	68		39	32	31	39			
					67	71	57	69			34	34	39			
					68	72	59				35	37	40			
						80	62				38	38	41			
							63				39	38	41			
							65				40	39	42			
							66				40	41	42			
							69				41	42	42			
							74				42	42	48			
											47	43	52			
											48	45	52			
											49	46	53			
											52	46	53			
											55	48	55			
											58	48	55			
											60	50	56			
												52	57			
												52	58			
												57	64			
												60	68			
												60	77			
												69				
Total																
N	-	3	2	6	10	11	16	9	1	8	23	29	28	6	3	2
Mean	-	33.67	38.50	42.17	55.00	58.91	55.13	50.33	32.00	28.62	38.91	41.10	46.00	55.50	65.00	53.50
Median	-	31.00	38.50	42.50	55.00	59.00	55.50	50.00	32.00	30.50	39.00	42.00	42.00	54.50	68.00	53.50
Eligible																
N	-	3	2	6	10	11	16	9	1	8	23	29	28	6	3	2
Mean	-	33.67	38.50	42.17	55.00	58.91	55.13	50.33	32.00	28.62	38.91	41.10	46.00	55.50	65.00	53.50
Median	-	31.00	38.50	42.50	55.00	59.00	55.50	50.00	32.00	30.50	39.00	42.00	42.00	54.50	68.00	53.50

EXHIBIT 20A (Continued)

	Female															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	26	17	10	12	37	37	40	24	-	27	16	17	13	51	40	41
		33	31	28	39	45	43	51		28	21	26	26	55	42	47
		41	38	29	40	47	45	54		29	24	28	29	63	59	52
			61	31	40	52	48	56		32	25	32	31	66	61	55
				31	41	55	54	57		36	31	33	33		62	62
				36	46	59	59	60		48	31	34	34		62	
				37	46	65	60	62			41	34	35		66	
				41	51	66	61	64			41	35	39		70	
				42	53	69	63	65			43	37	40			
				43	53	75	64	68			43	37	42			
				45	58		66	73			44	38	44			
				51	60		69				50	38	48			
				54	60		69				50	39	50			
				55	67		79				51	39	50			
											54	40	51			
											56	40	52			
											61	40	54			
												41	61			
												42	64			
												43	65			
												43	65			
												44	69			
												47				
												48				
												48				
												49				
												50				
												50				
												50				
												51				
												51				
												51				
												55				
												57				
Total																
N	1	3	4	14	14	10	14	11	-	6	17	34	22	4	8	5
Mean	26.00	30.33	35.00	38.21	49.36	57.00	58.57	57.64	-	33.33	40.12	41.38	45.23	58.75	57.75	51.40
Median	26.00	33.00	34.50	39.00	48.50	57.00	60.50	60.00	-	30.50	43.00	40.00	46.00	59.00	61.50	52.00
Eligible																
N	1	3	4	14	14	10	14	11	-	6	17	34	22	4	8	5
Mean	26.00	30.33	35.00	38.21	49.36	57.00	58.57	57.64	-	33.33	40.12	41.38	45.23	58.75	57.75	51.40
Median	26.00	33.00	34.50	39.00	48.50	57.00	60.50	60.00	-	30.50	43.00	40.00	46.00	59.00	61.50	52.00

EXHIBIT 20B - DISTRIBUTION OF PSI RAW SCORES FOR CHILDREN IN MEDIUM-TERM CENTERS, BY SEX, RACE, AND AGE

	Male															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	32	26 28*(1) 30 35 46* 47 51	21 24 26 32 35 36 38 39 39 39* 40 40 40 42 45* 46 47 48 49 49 49 50 51 51 52* 54 55 56 57 60* 60 61 64 66 66 70 71*	16 21 28 28 30 33 36* 36 37 37 39 41 44 45 45* 47 49 49 50 51 52 53* 53 55 55 55 61* 64 65 66 68* 68 68 69 72 72* 75* 77*	29 32 34 41* 42* 42 47 55 58 58 59 59 59* 60 61 61 62 63* 64 66* 68 69 72 78	36 46 49 54 54 58 61 61 62 62 65 65 65 66 70 70 70 71 72 74 74 75 76 78 78 84	41 47 50 58* 59* 61 67 73	32* 35	04 18 18 22 31 51	13 14 19 22 32 33 35 40* 44 44* 46* 50	16* 28 32* 33 34 34 37 39* 40 41 42* 46* 47 52* 54 57 65	31 31* 32* 33* 34 37 38 39 43 45* 46 51	39	59	49
Total																
N	-	1	7	37	38	24	26	8	2	6	12	17	12	1	1	1
Mean	-	32.00	37.57	47.78	50.26	55.79	65.23	57.00	33.50	24.00	32.67	41.00	38.33	39.00	59.00	49.00
Median	-	32.00	35.00	49.00	50.50	59.00	65.50	58.50	33.50	20.00	34.00	40.00	37.50	39.00	59.00	49.00
Eligible																
N	-	1	5	32	30	19	26	6	1	6	9	11	8	1	1	1
Mean	-	32.00	37.80	46.91	47.43	56.21	65.23	56.50	35.00	24.00	29.11	42.73	39.88	39.00	59.00	49.00
Median	-	32.00	35.00	48.50	49.00	59.00	65.50	55.50	35.00	20.00	32.00	40.00	38.50	39.00	59.00	49.00

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 17 weeks.

EXHIBIT 20B (Continued)

	Female															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	17 25 35 50	28 31 33 46 49 53 70	20 26 28 28 33 34 35*(1) 36 37 38 38 39 44 45 45* 46* 47* 48 52 54 55 56* 56 56 57 60* 64 64 66 69	15 19 28 32 33 39 40* 48 50 50 52 52 53* 54* 55 55* 56 57 57 58 59 59* 60 61* 61 65 66 72 72 73*	19 31 38 45 48 50 57 57 58 61 62 63 65 65 66 67 68 72 73 75	32 34 35 38 42 45 47 48 61 63 64 65 68 70 70 77 81	37 68 70 74 82 83 84	29*	20 24 25 30 34 37 41	30* 33 35 35 42	33 33 37* 38 39* 40 41 42 43 49 50 58 64	20* 29 36 38 44* 48* 51 52 55* 56 66*	43 63 67 69 71	76	54 65
Total																
N	-	4	7	30	30	20	17	7	1	7	5	13	11	5	1	2
Mean	-	31.75	44.29	45.87	51.70	57.00	55.29	71.14	29.00	30.14	35.00	43.62	45.00	62.60	76.00	59.50
Median	-	30.00	46.00	45.50	55.00	61.50	61.00	74.00	29.00	30.00	35.00	41.00	48.00	67.00	76.00	59.50
Eligible																
N	-	4	7	24	23	20	17	7	0	7	4	11	6	5	1	2
Mean	-	31.75	44.29	45.29	50.26	57.00	55.29	71.14	0.00	30.14	36.25	44.64	43.67	62.60	76.00	59.50
Median	-	30.00	46.00	44 50	55.00	61.50	61.00	74.00	0.00	30.00	35.00	42.00	44.50	67.00	76.00	59.50

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 17 weeks.

EXHIBIT 20C - DISTRIBUTION OF PSI RAW SCORES FOR CHILDREN IN LONG-TERM CENTERS, BY SEX. RACE, AND AGE

	Male							
	White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	06*(1)	-	16	24* 28 30* 35 39 40 40 42 43 66 74	29 32 43* 50 52 56* 67 69 69	47 72*	-	-
Total								
N	1	-	1	11	9	2	-	-
Mean	6.00	-	16.00	41.91	51.89	59.50	-	-
Median	6.00	-	16.00	40.00	52.00	59.50	-	-
Eligible								
N	0	-	1	9	7	1	-	-
Mean	0.00	-	16.00	45.22	52.57	47.00	-	-
Median	0.00	-	16.00	40.00	52.00	47.00	-	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 20C (Continued)

	Male							
	Non-White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	22*(1) 26	13 16 18 24 25* 27* 28* 28* 28* 28 33 33 41 45 47 47*	24 24 28 30* 32* 34* 34 36 37 41 43 46 48 51 53	19 19 22* 23 24 27 27 27 28 28 30 32 33* 35* 35 37 37* 38 39* 39* 39 42 42 43 43 43 44 45 45 45 45 46 46 46* 47 47 48 49 50 51 51* 51 52 53* 54* 54 55 56 59 62* 62 64 73	23 23* 26* 29 31 31* 38* 40 40 43 44 44 46 47 47 50 51 52 52 54 54* 54 55 56 57 57 58 59 59 60 70	34 41* 48 57	61	-
Total								
N	2	16	15	53	31	4	1	-
Mean	24.00	30.06	37.40	42.47	46.77	45.00	61.00	-
Median	24.00	28.00	36.00	44.00	50.00	44.50	61.00	-
Eligible								
N	1	10	12	42	26	3	1	-
Mean	26.00	29.80	38.75	42.38	49.15	46.33	61.00	-
Median	26.00	30.50	39.00	44.50	51.50	48.00	61.00	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 20C (Continued)

	Female							
	White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	33	36	30	24	54	-	-
		34	66	39	27			
				44	29*(1)			
					51			
					57			
					58*			
					63			
Total								
N	-	2	2	3	7	1	-	-
Mean	-	33.50	51.00	37.67	44.14	54.00	-	-
Median	-	33.50	51.00	39.00	51.00	54.00	-	-
Eligible								
N	-	2	2	3	5	1	-	-
Mean	-	33.50	51.00	37.67	44.40	54.00	-	-
Median	-	33.50	51.00	39.00	51.00	54.00	-	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 20C (Continued)

	Female							
	Non-White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	24 32 45	10 26 29*(1) 32 36 43*	22 24* 26* 29* 34 35* 36 37 38 38 39* 40* 42* 43 45 46 49 50* 54*	21 24* 28* 29 29 29 29 30 30* 31 31* 31 31 32* 33 33* 34 34* 36 37 37* 38* 39* 39 41 41 43 44 45 47* 47 49* 49 49 50 51* 51 51 52 53 53 55* 57* 58* 59 62 63* 66	16 29 32 32 35 35 36 38 39* 40 40 40 41 42 44* 44 47* 47* 48* 49 50 50 50* 50 51 52 52 52 52 53 55 56* 56* 57 58* 58 58 58 60 60 61* 61 64* 65 74*	51 59 60 64	64	-
Total								
N	3	6	19	48	45	4	1	-
Mean	33.67	29.33	38.26	41.69	48.60	58.50	64.00	-
Median	32.00	30.50	38.00	40.00	50.00	59.50	64.00	-
Eligible								
N	3	4	10	31	33	4	1	-
Mean	33.67	26.00	38.80	41.77	46.79	58.50	64.00	-
Median	32.00	29.00	38.00	41.00	50.00	59.50	64.00	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 21A - DISTRIBUTION OF BI RAW SCORES FOR CHILDREN IN SHORT-TERM CENTERS, BY SEX, RACE, AND AGE

	Male															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	117 135 168	146 165	124 133 143 147 161 182	95 108 110 114 117 132 141 157 161 170	111 119 127 129 137 151 161 163 164 175 184	123 127 142 142 151 152 158 158 160 164 166 167 174 179 183 184	107 110 121 144 146 147 148 178 191	145	101 109 111 118 123 140 140 152	94 105 111 120 122 132 132 138 142 145 146 149 150 151 154 159 159 160 162 164 172 175 183	69 102 104 105 106 106 107 108 117 119 119 120 122 123 125 130 132 136 142 149 156 157 168 168 173 175 175 180 183	93 106 107 115 117 121 122 122 123 123 124 126 131 133 135 137 137 137 139 140 141 142 159 165 172 175 181 184	116 120 140 147 151 181	138 165 173	133 164
Total																
N	-	3	2	6	10	11	16	9	1	8	23	29	28	6	3	2
Mean	-	140.00	155.50	148.33	130.50	147.36	158.13	143.36	145.00	124.25	144.57	133.65	135.96	142.50	158.67	148.50
Median	-	135.00	155.50	145.00	124.00	151.00	159.00	146.00	145.00	120.50	149.00	125.00	134.00	143.50	165.00	148.50
Eligible																
N	-	3	2	6	10	11	16	9	1	8	23	29	28	6	3	2
Mean	-	140.00	155.50	148.33	130.50	147.36	158.13	143.36	145.00	124.25	144.57	133.65	135.96	142.50	158.67	148.50
Median	-	135.00	155.50	145.00	124.00	151.00	159.00	146.00	145.00	120.50	149.00	125.00	134.00	143.50	165.00	148.50

EXHIBIT 21A (Continued)

	Female															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	143	123 135 171	147 148 170 188	100 104 117 120 139 139 140 149 151 151 153 156 160 197	105 115 128 129 137 140 144 145 146 148 152 154 160 160	105 120 126 141 145 149 155 158 166 177	112 116 131 137 138 139 143 144 145 154 154 163 164 173	134 135 140 140 145 146 150 154 157 157 163	-	107 117 140 164 172 185	111 114 122 125 128 130 144 147 147 149 155 162 164 174 178 188 193	117 124 130 130 132 133 134 135 137 138 139 142 142 143 143 148 151 153 154 155 155 158 160 160 164 166 167 168 168 169 169 170 172 183	92 118 120 121 121 124 126 127 131 139 142 144 145 147 152 159 164 164 167 182 184	120 156 159 171	107 132 138 139 147 149 158 166	104 105 147 154 156
Total																
N	1	3	4	14	14	10	14	11	-	6	17	34	22	4	8	5
Mean	143.00	143.00	163.25	141.14	140.21	144.20	143.79	147.36	-	147.50	148.89	150.26	142.59	151.50	142.00	133.20
Median	143.00	135.00	159.00	144.50	144.50	147.00	143.50	146.00	-	152.00	147.00	152.00	143.00	157.50	143.00	147.00
Eligible																
N	1	3	4	14	14	10	14	11	-	6	17	34	22	4	8	5
Mean	143.00	143.00	163.25	141.14	140.21	144.20	143.79	147.36	-	147.50	148.89	150.26	142.59	151.50	142.00	133.20
Median	143.00	135.00	159.00	144.50	144.50	147.00	143.50	146.00	-	152.00	147.00	152.00	143.00	157.50	143.00	147.00

EXHIBIT 21B - DISTRIBUTION OF BI RAW SCORES FOR CHILDREN IN MEDIUM-TERM CENTERS, BY SEX, RACE, AND AGE

	Male															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	166	97 103 105 111*(1) 115* 123 161	83* 100 103 106 114 115 123 123 125 129 129 130 130 130 138 138 140 145 146 147 152* 153 153 155 156* 158* 158 158 163 164 165 167 167 169 174 175 184*	107 115* 119 123 127 128* 132* 135 136 136 137 138 139* 142 142 145 150 151 152 153 154 154 155 155* 157 158 159 160 161 162 163 165 168 168 172* 174 175* 182*	106 113* 115 119 123 126 129* 132 132 135 135* 136 137 137* 141 144 145 154* 159 166 167 170 176 178	101 120 122 123 123 123 125 126 126 131 132 132 133 147 148 148 149 151 152 152 153 166 166 176 176 180	94 114 118 125 130 132* 137* 151	122 139*	102 104 108 114 150 177	72 125 128* 128 130* 130 132 133* 137 149 149 178	83* 110 112 120* 120* 123 124 124* 125 131 134* 134 144 156* 164 168 175	106 111 112 128* 138* 144 152 164 169* 173 183* 187	140	162	132
Total																
N	-	1	7	37	38	24	26	8	2	6	12	17	12	1	1	1
Mean	-	166.00	116.43	142.30	148.66	147.54	141.58	125.13	130.50	125.83	132.58	132.18	147.25	140.00	162.00	132.00
Median	-	166.00	111.00	146.00	152.50	136.50	140.00	128.50	130.50	111.00	131.00	125.00	148.00	140.00	162.00	132.00
Eligible																
N	-	1	5	32	30	19	26	6	1	6	9	11	8	1	1	1
Mean	-	166.00	117.80	141.63	148.37	151.21	141.58	122.00	122.00	125.83	133.33	137.27	143.63	140.00	162.00	132.00
Median	-	166.00	105.00	142.50	152.50	137.00	140.00	121.50	122.00	111.00	132.00	125.00	148.00	140.00	162.00	132.00

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 17 weeks.

EXHIBIT 21B (Continued)

	Female															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	84 119 125 171	112 120 135 152 152 164 185	108 108 109 111 113 116 122 122 126*(1) 132 134 135 135 135* 139 139* 143* 144 145 150* 151* 152 158 161 164 171 171 175 176 189	105 108 109 113 116* 119* 120 125 127 128 129 129 134 138 141 141 144* 145 145* 147 154 158 162 162 166 169* 172* 184* 193 194	96 109 115 117 122 123 135 136 137 139 139 140 145 150 155 164 165 167 179 182	95 107 113 114 116 118 121 124 131 141 145 150 156 160 161 171 182	97 130 143 147 152 171 187	136*	100 112 129 138 138 150 172	123 129 140 150* 184	97 102 108* 119 130 133 144 149 153 160 162* 182 183	101* 114* 120 121 122* 127 141 147* 147 154 184*	142 142 144 180 181	167	114 132
Total																
N	-	4	7	30	30	20	17	7	1	7	5	13	11	5	1	2
Mean	-	124.75	145.71	141.13	142.57	140.75	135.59	146.71	136.00	134.14	145.20	140.15	134.36	157.80	167.00	123.00
Median	-	122.00	152.00	139.00	141.00	139.00	131.00	147.00	136.00	138.00	140.00	144.00	127.00	144.00	167.00	123.00
Eligible																
N	-	4	7	24	23	20	17	7	0	7	4	11	6	5	1	2
Mean	-	124.75	145.71	141.25	140.35	140.75	135.59	146.71	0.00	134.14	144.00	141.09	135.00	157.80	167.00	123.00
Median	-	122.00	152.00	137.00	138.00	139.00	131.00	147.00	0.00	138.00	134.50	144.00	134.00	144.00	167.00	123.00

EXHIBIT 21C - DISTRIBUTION OF BI RAW SCORES FOR CHILDREN IN LONG-TERM CENTERS, BY SEX, RACE, AND AGE

	Male							
	White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	110*(1)	-	130	114*	112*	118*	-	-
				118	133*	149		
				130	137			
				139	138			
				143	147			
				145	148			
				148	155			
				160	155			
				173	158			
				173				
				176*				
Total								
N	1	-	1	11	9	2	-	-
Mean	110.00	-	130.00	147.18	142.56	133.50	-	-
Median	110.00	-	130.00	145.00	147.00	133.50	-	-
Eligible								
N	0	-	1	9	7	1	-	-
Mean	0.00	-	130.00	147.67	148.29	149.00	-	-
Median	0.00	-	130.00	145.00	148.00	149.00	-	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 21C (Continued)

	Male							
	Non-White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	173*(1) 176	96* 111 118* 124 126* 127* 131* 134 134* 135 144 147 148 158 160 167	82 101 118 133 134* 138 149 152* 153* 153 160 163 166 169 177	81 82* 94* 94 96 100 110 112 115 115 116 116 119 122 123 124 125 126 128 130 132 133 135* 135 136* 136 136 137* 139 141 143 149* 151* 151 153* 153 159* 161 161* 163 164 164 165 168 168 170* 175 177 179 182 182 185 185	119 126 127* 128 128* 129 130* 131 132 135 138 140 142 144 146* 147 154 155 156 156 159* 163 163 163 167 169 172 176 179 182 185	86* 123 131 155	168	-
Total								
N	2	16	15	53	31	4	1	-
Mean	174.50	135.00	143.20	139.55	149.71	123.75	168.00	-
Median	174.50	134.00	152.00	136.00	147.00	127.00	168.00	-
Eligible								
N	1	10	12	42	26	3	1	-
Mean	176.00	142.80	142.42	139.74	151.96	136.33	168.00	-
Median	176.00	145.50	151.00	135.50	154.50	131.00	168.00	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 21C (Continued)

	Female							
	White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	125 129	137 142	120 162 183	64 94* 122 153 157* 172 184	161	-	-
Total								
N	-	2	2	3	7	1	-	-
Mean	-	127.00	139.50	155.00	135.14	161.00	-	-
Median	-	127.00	139.50	162.00	153.00	161.00	-	-
Eligible								
N	-	2	2	3	5	1	-	-
Mean	-	127.00	139.50	155.00	139.00	161.00	-	-
Median	-	127.00	139.50	162.00	153.00	161.00	-	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 21C (Continued)

	Female							
	Non-White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	96 132 152	117 142*(1) 144* 144 155 168	110* 124 124 127 140* 141 142* 143 146* 148 151 151* 152 154 157 158* 167* 168* 173*	104 108* 111 112 114 116 117 119 121 121* 124 127 134 134 134 135* 136* 136* 137 137 138* 141 141* 143 144 148 150 153* 155* 155 156 159* 159* 160 161 163* 164* 166* 167 168 168 170 171* 177* 178 184* 188 194	99 102 108 110 114 129 130 132* 134 137 140 140 140 141 147 147 150 151* 151* 152 153* 155 156 156 157* 157 158* 158* 158* 159 161* 165 167 170 172 173 174* 174* 176* 178 184 185 187 191 191	119 137 157 177	175	-
Total								
N	3	6	19	48	45	4	1	-
Mean	126.67	145.00	146.11	145.80	152.64	147.50	175.00	-
Median	132.00	144.00	148.00	143.50	156.00	147.00	175.00	-
Eligible								
N	3	4	10	31	33	4	1	-
Mean	126.67	146.00	142.00	142.97	149.94	147.50	175.00	-
Median	132.00	149.50	145.50	141.00	151.00	147.00	175.00	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 22A - DISTRIBUTION OF VSMS RAW SCORES FOR CHILDREN IN SHORT-TERM CENTERS, BY SEX, RACE, AND AGE

	Male															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	49.0	56.0	48.5	50.0	47.0	58.5	57.0	46.0	41.5	46.5	43.0	39.0	57.0	60.0	62.5
		50.0	59.0	53.5	55.0	57.5	59.0	57.0		45.0	49.5	48.0	48.0	58.0	61.0	71.5
		51.5		55.0	55.5	59.0	59.5	58.0		46.0	52.0	49.5	49.0	58.0	69.5	
				57.5	56.5	59.5	59.5	60.0		51.0	52.0	50.0	49.5	59.0		
				58.0	58.0	60.0	60.0	60.5		52.0	52.0	50.0	52.0	59.5		
				59.5	58.5	60.5	60.5	61.0		54.0	52.0	51.0	54.0	69.5		
					59.0	60.5	60.5	61.5		55.0	52.5	52.5	55.0			
					59.5	61.0	60.5	61.5		56.5	53.0	53.0	55.5			
					60.5	62.0	61.0	63.0			53.5	53.0	56.0			
					61.0	62.0	61.0				53.5	53.5	57.0			
						64.5	61.0				56.0	53.5	57.0			
							61.5				56.0	54.0	57.0			
							62.0				57.5	55.5	57.0			
							63.0				58.0	55.5	57.0			
							64.5				58.5	56.0	57.0			
							65.0				58.5	57.0	57.5			
											59.0	57.0	58.0			
											59.0	57.5	58.0			
											60.5	58.0	58.0			
											60.5	58.0	58.0			
											61.0	59.0	58.5			
											69.5	59.5	59.0			
												59.5	60.0			
												60.0	64.5			
												61.5	65.0			
												62.5	65.0			
												63.0	66.0			
												64.0	69.5			
												64.5				
Total																
N	-	3	2	6	10	11	16	9	1	8	22	29	28	6	3	2
Mean	-	50.17	57.50	55.33	57.35	59.41	61.06	59.94	46.00	50.13	55.98	55.83	57.04	60.17	63.50	67.00
Median	-	50.00	57.50	56.25	58.25	60.50	60.75	60.50	46.00	51.50	56.00	56.00	57.00	58.50	61.00	67.00
Eligible																
N	-	3	2	6	10	11	16	9	1	8	22	29	28	6	3	2
Mean	-	50.17	57.50	55.33	57.35	59.41	61.06	59.94	46.00	50.13	55.98	55.83	57.04	60.17	63.50	67.00
Median	-	50.00	57.50	56.25	58.25	60.50	60.75	60.50	46.00	51.50	56.00	56.00	57.00	58.50	61.00	67.00

EXHIBIT 22A (Continued)

	Female															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	47.0	42.0 52.5 55.5	53.0 54.0 54.5 60.0	52.0 53.5 54.0 54.0 55.0 55.0 56.0 56.5 57.0 57.0 59.0 59.5 59.5 60.0	52.0 53.5 54.0 55.5 55.5 56.0 56.0 58.0 58.5 60.0 60.0 60.5 61.0 63.0	50.5 52.0 54.0 56.0 57.5 59.5 60.0 60.0 61.0 61.0	53.5 55.5 57.5 58.5 58.5 59.0 60.0 60.0 60.5 60.5 60.5 61.0 61.5 67.5	55.0 56.0 56.0 57.5 58.0 58.5 59.5 59.5 60.0 60.5 60.5	-	41.5 53.0 56.0 57.0 58.0 58.5	45.0 48.5 48.5 50.5 51.5 52.5 53.0 55.0 56.5 57.0 57.0 57.0 57.5 57.5 58.0 58.0 60.5 61.0	31.0 47.0 48.5 49.5 49.5 51.5 52.5 53.5 54.0 54.5 55.0 55.0 56.5 56.5 56.5 56.5 57.0 57.0 57.0 57.5 57.5 57.5 57.5 58.0 58.0 58.5 58.5 59.0 61.0 61.5 62.0 62.0 63.0 64.5	51.5 52.0 52.5 52.5 55.0 55.0 55.0 55.5 56.0 56.5 56.5 57.0 58.0 59.0 60.0 61.5 61.5 64.0 64.5 66.0 68.5 71.0	53.0 56.0 58.0 64.5	59.0 61.0 61.0 62.0 62.5 63.5 64.0 66.5	55.0 60.5 62.0 65.0 69.0
Total																
N	1	3	4	14	14	10	14	11	-	6	17	34	22	4	8	5
Mean	47.00	50.00	55.38	56.29	57.39	57.15	58.86	58.27	-	54.00	54.56	55.72	58.59	57.88	62.44	62.30
Median	47.00	52.50	54.25	56.25	58.00	58.50	60.00	58.50	-	56.50	56.50	57.00	56.75	57.00	62.25	62.00
Eligible																
N	1	3	4	14	14	10	14	11	-	6	17	34	22	4	8	5
Mean	47.00	50.00	55.38	56.29	57.39	57.15	58.86	58.27	-	54.00	54.56	55.72	58.59	57.88	62.44	62.30
Median	47.00	52.50	54.25	56.25	58.00	58.50	60.00	58.50	-	56.50	56.50	57.00	56.75	57.00	62.25	62.00

EXHIBIT 22B - DISTRIBUTION OF VSMS RAW SCORES FOR CHILDREN IN MEDIUM-TERM CENTERS, BY SEX, RACE, AND AGE

	Male															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	53.5	47.0*(1)	35.0	39.5	47.0	55.5	55.5	51.5	44.5	44.5	28.5	45.5	55.5	59.5	56.0
			48.5	48.5	42.0	50.0*	56.0	57.0	59.0*	45.0	45.5	47.0	47.5			
			49.0	49.0	49.0	51.0	58.0	58.0		48.0	47.5	48.0*	51.5			
			51.5	51.5	50.0	53.0	58.0	61.5		48.0	51.0	49.5*	53.0			
			53.0*	52.0	50.0	54.5	58.0	62.0		59.0	52.5	49.5*	57.5			
			53.0	52.5*	51.5	55.5*	59.0	62.0*		59.0	54.0*	52.0	57.5			
			54.0	52.5	54.0	56.0	59.0	62.0			55.0	53.5	57.5			
				53.0	54.0	57.0	60.0	65.0*			55.5	54.0	59.5*			
				53.5	54.5	57.5	60.0				55.5*	54.5	59.5*			
				54.0	54.5*	58.0	60.0				57.5	57.0	60.0			
				55.0	55.0	59.0*	60.5				58.0	58.0	62.5*			
				55.0	55.5	59.5	61.0				58.0*	59.0	62.5*			
				55.0	55.5	59.5	61.0					59.5				
				56.0	55.5	59.5	62.0					62.5				
				56.0	57.0	59.5	62.5					62.5*				
				56.0	57.0	59.5	63.5					64.0*				
				56.5	57.5	60.0	63.5					64.5*				
				57.0	58.0	60.0	64.0									
				57.0	58.0	61.0	64.0									
				57.0	58.5	61.5*	64.0									
				57.0	59.0	61.5*	65.0									
				57.5	59.0*	64.0	65.0									
				58.0	60.5	64.5	65.0									
				58.0	60.5*	66.5	65.5									
				59.0	61.0		65.5									
				59.0	61.0		67.0									
				60.0*	61.5											
				60.0	62.0											
				60.5	62.0											
				61.5	63.0*											
				62.0*	63.5											
				62.0*	65.0*											
				63.0	64.5											
				64.0	65.0											
				65.5*	65.5											
				66.5	67.5*											
				68.5	70.5*											
					70.5*											
Total																
N	-	1	7	37	38	24	26	8	2	6	12	17	12	1	1	1
Mean	-	53.50	50.90	56.87	58.11	58.13	61.63	60.38	55.25	50.58	52.88	54.85	56.17	55.50	59.50	56.00
Median	-	53.50	51.50	57.00	58.25	59.50	61.50	61.75	55.25	48.00	54.50	54.50	57.50	55.50	59.50	56.00
Eligible																
N	-	1	5	32	30	19	26	6	1	6	9	11	8	1	1	1
Mean	-	53.50	51.20	56.31	56.58	58.29	61.63	59.33	51.50	50.58	51.88	53.23	53.75	55.50	59.50	56.00
Median	-	53.50	51.50	56.75	57.25	59.50	61.50	59.75	51.50	48.00	52.50	54.50	55.25	55.50	59.50	56.00

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 17 weeks.

EXHIBIT 22B (Continued)

	Female															
	White							Non-White								
	Age Range							Age Range								
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	41.0 50.0 50.5 55.5	50.5 52.0 53.5 54.0 55.0 57.5 65.5	44.5 46.5*(1) 46.5 47.0 49.0 49.5 49.5 51.0 51.5 52.5 52.5 53.0* 53.5 55.0* 55.0 55.0 55.5 56.0 56.5* 57.0 57.0 59.5 59.5 61.5* 61.5 62.0 62.5* 63.0 66.0 66.5	46.0 49.0 49.5* 53.0 53.0 54.0 55.0 55.0 55.0* 55.0 55.5 55.5 56.0 56.5 57.0 57.0 57.5 58.0 58.5 58.5 59.0* 59.0* 60.0 60.5 61.5* 62.0 62.0 63.0* 63.5* 63.5	47.0 54.0 56.0 56.5 57.5 58.5 58.5 59.5 60.0 60.5 60.5 61.0 61.5 62.0 62.5 63.0 64.0 65.0 66.0 68.0	46.0 53.5 54.5 58.0 58.0 59.0 59.5 59.5 60.5 61.0 61.5 61.5 63.0 63.0 63.5 65.5 66.5	53.5 60.0 62.5 64.0 66.0 68.0 69.5	47.0*	47.5 48.5 50.0 53.0 55.0 56.0 59.5	45.5 55.0 57.5 59.5* 62.0	49.5 52.0 54.5 56.0 56.0* 56.5 58.0* 59.0 59.0 60.0 60.0 61.5 64.0	32.0* 51.5 54.0 58.0* 59.0* 59.0 60.0* 61.0 62.5 63.0* 66.0	57.5 57.5 59.5 63.5 64.0	60.5	59.5 59.5
Total																
N	-	4	7	30	30	20	17	7	1	7	5	13	11	5	1	2
Mean	-	49.25	55.43	55.18	56.95	60.08	59.65	63.36	47.00	52.79	55.90	57.38	56.91	60.40	60.50	59.50
Median	-	50.25	54.00	55.00	57.00	60.50	59.50	64.00	47.00	53.00	57.50	58.00	59.00	59.50	60.50	59.50
Eligible																
N	-	4	7	24	23	20	17	7	0	7	4	11	6	5	1	2
Mean	-	49.25	55.43	55.02	56.43	60.08	59.65	63.36	0.00	52.79	55.00	57.45	59.00	60.40	60.50	59.50
Median	-	50.25	54.00	55.00	56.50	60.50	59.50	64.00	0.00	53.00	56.25	59.00	60.00	59.50	60.50	59.50

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 17 weeks.

EXHIBIT 22C - DISTRIBUTION OF VSMS RAW SCORES FOR CHILDREN IN LONG-TERM CENTERS, BY SEX, RACE, AND AGE

	Male							
	White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	40.5*(1)	-	54.5	47.5* 52.0 53.5 54.0 55.0 55.0 57.0 58.5 60.5 62.0* 63.0	52.5 54.0 57.0 58.0* 58.5 59.0* 60.0 62.0 62.5	53.0 71.0*	-	-
Total								
N	1	-	1	11	9	2	-	-
Mean	40.50	-	54.50	56.18	58.17	62.00	-	-
Median	40.50	-	54.50	55.00	58.50	62.00	-	-
Eligible								
N	0	-	1	9	7	1	-	-
Mean	0.00	-	54.50	56.50	58.07	53.00	-	-
Median	0.00	-	54.50	55.00	58.50	53.00	-	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 22C (Continued)

	Male							
	Non-White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	45.5*(1)	36.5	45.0	38.5	47.0	55.0*	63.0	-
	59.0	40.0	49.5	46.5*	52.0	56.5		
		45.0	50.0	47.0	52.0	57.0		
		45.0	50.0*	48.5	52.0	59.0		
		45.0*	50.5	49.0	52.5			
		46.0*	52.5*	49.5	53.5			
		48.5*	53.5	49.5	54.5			
		49.0*	53.5	50.0	56.0			
		50.5	55.5	50.5	56.0			
		51.0	56.5	51.0	56.5			
		51.0	57.5	51.5*	56.5			
		51.0*	58.0	52.0	56.5*			
		52.5	58.0	52.5	57.0			
		54.5	58.0	52.5	57.0*			
		56.0	61.0*	53.0	58.5			
		59.0*		53.0	58.5			
				53.0	58.5			
				53.5	58.5*			
				54.0	59.5*			
				55.0	60.5			
				55.0	60.5			
				55.5	61.0			
				56.0	61.5			
				56.0	61.5			
				56.0	62.0*			
				57.0*	63.0			
				57.5	63.5			
				57.5	64.5			
				57.5*	65.5			
				58.0*	66.5			
				58.0	70.5			
				58.0				
				58.0				
				58.5				
				58.5				
				59.0*				
				59.0				
				59.0*				
				59.5				
				59.5				
				59.5*				
				60.0				
				60.0				
				60.0*				
				61.0				
				61.0				
				61.0*				
				61.0				
				61.0				
				62.0				
				62.5				
				63.0*				
				66.0				
Total								
N	2	16	15	53	31	4	1	
Mean	52.25	48.78	53.93	55.87	58.48	56.88	63.00	-
Median	52.25	49.75	53.50	57.50	58.50	56.75	63.00	-
Eligible								
N	1	10	12	42	26	3	1	-
Mean	59.00	48.20	53.79	55.57	58.44	57.50	63.00	-
Median	59.00	50.75	54.50	56.00	58.50	57.00	63.00	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 22C (Continued)

	Female							
	White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	48.5	54.0	50.5	52.5*(1)	67.5	-	-
		53.5	54.0	55.0	58.5			
				66.0	59.5			
					61.0*			
					62.0			
					62.0			
					67.5			
Total								
N	-	2	2	3	7	1	-	-
Mean	-	51.00	54.00	57.17	60.43	67.50	-	-
Median	-	51.00	54.00	55.00	61.00	67.50	-	-
Eligible								
N	-	2	2	3	5	1	-	-
Mean	-	51.00	54.00	57.17	61.90	67.50	-	-
Median	-	51.00	54.00	55.00	62.00	67.50	-	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT 22C (Continued)

	Female							
	Non-White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	49.0	42.5	46.0	45.5*	46.5	58.5	67.0	-
	51.5	46.5	46.0*	48.5	47.5	59.0		
	58.0	52.5*(1)	46.5*	48.5	50.5*	60.0		
		52.5	51.0*	51.5	51.5	62.0		
		56.0*	51.5	52.5	52.0			
		61.5	51.5*	53.5	54.0*			
			52.5	53.5	55.0*			
			53.0	54.0	55.5			
			54.0	54.0	56.0*			
			54.0	54.5	56.5*			
			54.0*	55.0	56.5*			
			55.0*	55.5	56.5			
			55.5	55.5	57.0			
			56.5*	55.5	58.0			
			58.0*	55.5*	58.0			
			58.0	55.5*	58.0			
			58.0	55.5	58.5			
			58.0*	56.5	58.5			
			59.0	56.5	58.5			
				56.5	58.5			
				56.5*	59.0			
				56.5*	59.5			
				56.5*	59.5			
				57.0	59.5			
				57.0	59.5			
				57.0*	59.5			
				57.5*	60.0			
				57.5*	60.0			
				57.5	60.0			
				58.0	60.0			
				58.0	60.0			
				58.0	60.0			
				58.5*	60.5			
				58.5*	61.0			
				58.5	61.5			
				59.0	61.5			
				59.0	61.5*			
				60.0	62.0			
				60.5*	62.0			
				60.5*	62.0*			
				60.5*	63.0			
				61.0	65.5*			
				62.0*	67.0*			
				63.0	69.0*			
				64.0	69.5*			
				65.0				
				65.0*				
				68.0				
Total								
N	3	6	19	48	45	4	1	-
Mean	52.83	51.92	53.58	57.15	58.79	59.88	67.00	-
Median	51.50	52.50	54.00	57.00	59.50	59.50	67.00	-
Eligible								
N	3	4	10	31	33	4	1	-
Mean	52.83	50.75	54.15	56.65	58.47	59.88	67.00	-
Median	51.50	49.50	54.00	56.50	59.50	59.50	67.00	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

the total number of children tested and included in PRC's data tabulations and analyses) and "eligible-only" children (N = 831). Scores of children tested in Spanish are not separately identified in the arrays.

Exhibits 23 through 26 present summary tables of unadjusted mean raw scores on the PPVT, PSI, BI, and VSMS for the subset of eligible children whose tests were administered in Spanish.

For the PPVT, PSI, BI, and VSMS, Exhibits 27 through 30 present, for S, M, and L eligible children only (N = 831), summaries of unadjusted raw score means, standard errors of those means, medians, and appropriate N's. These summary tables were prepared from the unit raw score distributions in Exhibits 19A through 22C. Medians were determined only when the sample size in a subclassification was less than 30. For the VSMS and the BI, standard errors of the means are presented for S, M, and L total means only.

Exhibits 31 and 32 present PSI subtest raw score unadjusted means for (1) all children, by age and race (Exhibit 31), and (2) eligible children only, by age and race (Exhibit 32). Exhibits 33 and 34 show the percentages of children answering the item correctly for each subtest for the same groups. These percentages indicate the relative difficulty of the different subtests. For example, in Exhibits 33 and 34 it is clear that subtest 2 was generally more difficult for children at each age level, regardless of race.

Subtest names[1] are:

- Subtest 1: Personal-Social Responsiveness
- Subtest 2: Associative Vocabulary
- Subtest 3: Numerical Concept Activation
- Subtest 4: Sensory Concept Activation

Finally, an item difficulty breakdown for the PSI subtests is given in Exhibit 35. Here, the percentage of eligible children (for

---

[1] Test names are those provided by the authors of the tests. See Caldwell, Bettye and Soule, Donald. The Pre-School Inventory (Unpublished Head Start Study, OEO Contract No. 514).

EXHIBIT 23 - PPVT SUMMARY OF RAW SCORE MEANS[1] AND SAMPLE SIZES FOR SHORT-, MEDIUM-, AND LONG-TERM ELIGIBLE CHILDREN, BY LANGUAGE USED IN TESTING, RACE, AND AGE

Age	Program Duration	White, Tested in Spanish	N	White, Tested in English	N	Non-White, Tested in English	N	Total, Tested in English	N	Total, Spanish and English	N
3	Short	-	-	29.14	7	28.67	15	28.82	22	28.82	22
	Medium	-	-	25.00	5	19.50	14	20.94	19	20.94	19
	Long	-	-	37.00	2	27.28	18	28.25	20	28.25	20
4	Short	23.67	3	37.21	24	36.31	102	36.48	126	36.19	129
	Medium	23.66	3	41.15	65	32.85	35	38.25	100	37.82	103
	Long	30.75	4	35.00	11	36.58	95	36.42	106	36.21	110
5	Short	42.00	8	45.41	37	39.83	60	42.72	97	42.65	105
	Medium	37.00	1	45.22	91	37.25	20	43.78	111	43.72	112
	Long	19.50	2	45.75	12	41.09	66	41.81	78	41.25	80
6	Short	43.63	43	44.14	7	43.67	18	44.40	25	43.76	68
	Medium	48.66	3	51.35	54	39.50	4	50.53	58	50.44	61
	Long	-	-	-	-	49.50	2	49.50	2	49.50	2
Total	Short	42.28	54	41.47	75	37.49	195	38.75	270	39.21	324
	Medium	36.29	7	44.98	215	31.86	73	41.65	288	41.53	295
	Long	27.00	6	41.60	25	37.44	181	37.95	206	37.49	212

Note: (1) Unadjusted means.

EXHIBIT 24 - PSI SUMMARY OF RAW SCORE MEANS(1) AND SAMPLE SIZES FOR SHORT-, MEDIUM-, AND LONG-TERM ELIGIBLE CHILDREN, BY LANGUAGE USED IN TESTING, RACE, AND AGE

Age	Program Duration	White, Tested in Spanish	N	White, Tested in English	N	Non-White, Tested in English	N	Total, Tested in English	N	Total, Spanish and English	N
3	Short	-	-	32.00	7	30.67	15	31.09	22	31.09	22
	Medium	-	-	33.00	5	28.07	14	29.37	19	29.37	19
	Long	-	-	32.00	2	29.50	18	29.75	20	29.75	20
4	Short	28.67	3	40.54	24	42.54	102	40.57	126	41.84	129
	Medium	35.33	3	45.92	65	38.86	35	43.45	100	42.92	103
	Long	33.25	4	45.64	11	41.11	95	41.58	106	41.27	110
5	Short	51.37	8	56.05	37	47.50	60	50.76	97	50.81	105
	Medium	27.00	1	52.38	91	46.25	20	51.28	111	51.06	112
	Long	24.50	2	45.33	12	48.97	66	48.41	78	48.00	80
6	Short	54.28	43	63.00	7	57.28	18	58.88	25	55.97	68
	Medium	55.33	3	62.74	54	55.75	4	62.26	58	61.92	61
	Long	-	-	-	-	62.00	2	62.00	2	62.00	2
Total	Short	52.43	54	49.49	75	43.49	195	45.16	270	46.37	324
	Medium	42.71	7	55.14	215	39.74	73	51.24	288	51.03	295
	Long	30.33	6	44.40	25	43.05	181	43.21	206	42.85	212

Note: (1) Unadjusted means.

EXHIBIT 25 - BI SUMMARY OF RAW SCORE MEANS[1] AND SAMPLE SIZES FOR SHORT-, MEDIUM-, AND LONG-TERM ELIGIBLE CHILDREN, BY LANGUAGE USED IN TESTING, RACE, AND AGE

Age	Program Duration	White, Tested in Spanish	N	White, Tested in English	N	Non-White, Tested in English	N	Total, Tested in English	N	Total, Spanish and English	N
3	Short	-	-	142.00	7	135.33	15	137.45	22	137.45	22
	Medium	-	-	133.00	5	129.50	14	130.42	19	130.42	19
	Long	-	-	127.00	2	142.56	18	141.00	20	141.00	20
4	Short	148.67	3	147.45	24	143.67	102	144.38	126	144.48	129
	Medium	125.33	3	140.85	65	137.86	35	139.80	100	139.77	103
	Long	159.50	4	142.45	11	144.32	95	141.34	106	142.00	110
5	Short	147.00	8	139.84	37	140.17	60	140.04	97	140.57	105
	Medium	147.00	1	143.65	91	143.25	20	143.58	111	143.65	112
	Long	-	2	150.33	12	148.82	66	149.05	78	148.44	80
6	Short	149.84	43	137.29	7	138.67	18	142.52	25	147.29	68
	Medium	154.33	3	138.20	54	134.50	4	137.95	58	138.80	61
	Long	-	-	-	-	172.00	2	172.00	2	172.00	2
Total	Short	149.35	54	142.71	75	141.85	195	142.08	270	143.30	324
	Medium	141.42	7	141.47	215	137.55	73	140.47	288	140.49	295
	Long	147.83	6	145.00	25	144.46	181	144.52	206	144.62	212

Note: (1) Unadjusted means.

EXHIBIT 26 - VSMS SUMMARY OF RAW SCORE MEANS[1] AND SAMPLE SIZES FOR SHORT-, MEDIUM-, AND LONG-TERM ELIGIBLE CHILDREN, BY LANGUAGE USED IN TESTING, RACE, AND AGE

Age	Program Duration	White, Tested in Spanish	N	White, Tested in English	N	Non-White, Tested in English	N	Total, Tested in English	N	Total, Spanish and English	N
3	Short	-	-	49.14	7	51.67	15	50.86	22	50.86	22
	Medium	-	-	51.00	5	52.00	14	51.74	19	51.74	19
	Long	-	-	49.50	2	50.61	18	50.50	20	50.50	20
4	Short	53.67	3	56.58	24	55.61	102	55.79	126	55.74	129
	Medium	53.67	3	55.62	65	54.57	35	55.25	100	55.20	103
	Long	59.50	4	55.18	11	55.63	95	55.58	106	55.73	110
5	Short	59.50	8	58.35	37	58.08	60	53.55	97	54.00	105
	Medium	52.00	1	58.15	91	57.00	20	57.95	111	57.89	112
	Long	62.00	2	59.08	12	58.74	66	58.79	78	58.88	80
6	Short	60.18	43	62.43	7	63.39	18	63.12	25	61.26	68
	Medium	57.67	3	61.07	54	60.75	4	61.05	58	61.09	61
	Long	-	-	-	-	64.50	2	64.50	2	64.50	2
Total	Short	59.72	54	57.30	75	56.77	195	56.92	270	57.38	324
	Medium	57.00	7	57.95	215	55.08	73	57.23	288	57.22	295
	Long	60.33	6	56.50	25	56.36	181	56.39	206	56.21	212

Note: (1) Unadjusted means.

EXHIBIT 27 - PPVT SUMMARY OF UNADJUSTED RAW SCORE MEANS, STANDARD ERRORS OF MEANS, MEDIANS, AND SAMPLE SIZES FOR ELIGIBLE CHILDREN IN SHORT-, MEDIUM-, AND LONG-TERM CENTERS

Sex	Race	Age	Short				Medium				Long			
			Mean	σ_M(1)	Median	N	Mean	σ_M	Median	N	Mean	σ_M	Median	N
Male	White	3	31.00	2.37	32.00	3	23.00	-	23.00	1	-	-	-	0
		4	36.75	3.89	37.50	8	41.95	1.71	**	37	38.00	5.94	39.50	10
		5	47.57	2.55	49.00	21	45.86	1.70	**	49	45.75	4.90	48.50	8
		6	44.76	1.19	44.00	25	52.46	3.69	**	32	-	-	-	0
	Non-White	3	26.44	5.53	22.00	9	19.28	5.20	18.00	7	29.00	5.89	26.00	11
		4	36.34	1.71	**(2)	52	32.60	2.40	31.50	20	36.70	1.87	**	54
		5	39.53	1.28	**	34	36.22	3.13	36.00	9	43.14	1.08	43.00	29
		6	46.60	4.27	48.00	5	43.50	0.50	43.50	2	52.00	-	52.00	1
Female	White	3	28.50	2.46	29.50	4	26.75	3.80	28.00	4	36.50	5.40	36.50	2
		4	34.56	3.29	35.50	18	38.00	1.64	**	31	32.60	1.16	30.00	5
		5	42.13	1.78	43.00	24	43.37	1.96	**	43	36.33	8.17	41.00	6
		6	43.56	2.35	43.00	25	48.67	2.82	51.00	24	-	-	-	0
	Non-White	3	33.00	1.56	31.00	6	20.71	1.97	20.00	7	27.00	4.10	28.00	7
		4	36.39	1.24	**	51	32.80	3.08	29.00	15	36.63	0.65	**	41
		5	39.54	2.10	40.50	26	39.00	4.02	42.00	11	39.24	1.60	**	37
		6	42.54	6.56	41.00	13	44.67	9.80	46.00	3	47.00	-	47.00	1
	Subtotal, by Age	3	29.23	NC(3)	30.00	22	21.57	NC	20.00	19	29.05	NC	28.50	20
		4	36.13	NC	**	129	24.05	NC	**	103	36.61	NC	**	110
		5	41.13	NC	**	105	43.46	NC	**	112	41.09	NC	**	80
		6	44.02	NC	**	68	50.30	NC	**	61	49.50	NC	49.50	2
	Total	All ages	38.94	0.28	**	324	36.68	1.60	**	295	37.71	1.45	**	212

Notes: (1) Standard error of the mean.

(2) The median was not obtained when N was more than 30.

(3) Not computed.

EXHIBIT 28 - PSI SUMMARY OF UNADJUSTED RAW SCORE MEANS, STANDARD ERRORS OF MEANS, MEDIANS, AND SAMPLE SIZES FOR ELIGIBLE CHILDREN IN SHORT-, MEDIUM-, AND LONG-TERM CENTERS

Sex	Race	Age	Short				Medium				Long			
			Mean	σ_M(1)	Median	N	Mean	σ_M	Median	N	Mean	σ_M	Median	N
Male	White	3	33.67	3.06	31.00	3	32.00	-	32.00	1	-	-	-	0
		4	41.25	4.89	41.50	8	45.67	2.11	**	37	42.30	5.03	40.00	10
		5	57.29	2.92	55.00	21	51.32	2.36	**	49	51.88	6.24	51.00	8
		6	53.40	2.57	53.00	25	63.59	8.80	**	32	-	-	-	0
	Non-White	3	29.00	4.08	32.00	9	25.57	7.46	22.00	7	29.10	4.95	28.00	11
		4	40.13	1.94	**(2)	52	36.60	2.80	34.50	20	41.57	2.13	**	54
		5	47.68	1.88	**	34	39.78	2.54	39.00	9	48.86	2.03	51.00	29
		6	60.40	11.87	68.00	5	54.00	5.00	54.00	2	61.00	-	61.00	1
Female	White	3	29.25	8.14	29.50	4	31.75	10.58	31.75	4	33.50	0.50	33.50	2
		4	37.50	4.04	37.50	18	45.06	3.00	**	31	43.00	6.17	39.00	5
		5	52.54	3.13	51.50	24	53.39	3.08	**	43	46.00	9.22	52.50	6
		6	58.16	3.27	60.00	25	59.92	4.80	64.50	24	-	-	-	0
	Non-White	3	33.33	4.30	30.50	6	30.14	3.60	30.00	7	29.29	4.13	32.00	7
		4	40.96	1.61	**	51	42.40	2.80	41.00	15	41.05	1.72	**	41
		5	47.27	2.80	50.00	26	52.27	4.98	52.00	11	48.03	5.98	**	37
		6	55.31	2.24	59.00	13	65.00	2.01	65.00	3	64.00	-	64.00	1
	Subtotal, by Age	3	30.86	NC(3)	31.50	22	28.89	NC	25.00	19	29.60	NC	32.00	20
		4	40.16	NC	**	129	43.25	NC	**	103	41.51	NC	**	110
		5	50.56	NC	**	105	51.29	NC	**	112	48.56	NC	**	80
		6	56.03	NC	**	68	61.90	NC	**	61	62.50	NC	62.50	2
	Total	All ages	46.09	1.52	**	324	49.23	2.21	**	295	43.25	1.23	**	212

Notes: (1) Standard error of the mean.

(2) The median was not obtained when N was more than 30.

(3) Not computed.

EXHIBIT 29 - VSMS SUMMARY OF UNADJUSTED RAW SCORE MEANS, STANDARD ERRORS OF MEANS, MEDIANS, AND SAMPLE SIZES FOR ELIGIBLE CHILDREN IN SHORT-, MEDIUM-, AND LONG-TERM CENTERS

Sex	Race	Age	Short				Medium				Long			
			Mean	σ_M (1)	Median	N	Mean	σ_M	Median	N	Mean	σ_M	Median	N
Male	White	3	50.17	NC(3)	50.00	3	53.50	NC	53.50	1	-	-	-	0
		4	55.87	NC	56.75	8	55.62	NC	**	37	56.30	NC	55.00	10
		5	58.43	NC	59.50	21	57.24	NC	**	49	57.44	NC	57.75	8
		6	60.66	NC	60.50	25	61.20	NC	**	32	-	-	-	0
	Non-White	3	49.69	NC	51.00	9	50.71	NC	48.00	7	49.18	NC	51.00	11
		4	55.90	NC	**(2)	52	52.62	NC	54.25	20	55.18	NC	**	54
		5	57.59	NC	**	34	53.94	NC	55.50	9	58.34	NC	58.50	29
		6	64.90	NC	62.50	5	57.75	NC	57.75	2	63.00	NC	63.00	1
Female	White	3	49.25	NC	49.75	4	49.25	NC	50.25	4	51.00	NC	51.00	2
		4	56.09	NC	55.50	18	55.11	NC	**	31	55.90	NC	54.00	5
		5	57.29	NC	57.75	24	58.13	NC	**	43	62.83	NC	62.00	6
		6	58.60	NC	59.50	25	60.73	NC	61.25	24	-	-	-	0
	Non-White	3	54.00	NC	56.50	6	52.79	NC	53.00	7	51.64	NC	51.50	7
		4	49.76	NC	**	51	56.80	NC	57.50	15	56.04	NC	**	41
		5	58.48	NC	56.75	26	59.64	NC	59.50	11	58.62	NC	**	37
		6	62.39	NC	62.00	13	59.83	NC	59.50	3	67.00	NC	67.00	1
	Subtotal, by Age	3	50.85	NC	51.75	22	51.31	NC	50.00	19	50.20	NC	51.00	20
		4	53.49	NC	**	129	55.06	NC	**	103	55.63	NC	**	110
		5	57.91	NC	**	105	57.55	NC	**	112	46.22	NC	**	80
		6	60.54	NC	**	68	60.83	NC	**	61	65.00	NC	65.00	2
	Total	All ages	56.23	0.73	**	324	56.96	1.90	**	295	51.66	1.89	**	212

Notes: (1) Standard error of the mean, computed only for total, all ages.

(2) The median was not obtained when N was more than 30.

(3) Not computed.

EXHIBIT 30 - BI SUMMARY OF UNADJUSTED RAW SCORE MEANS, STANDARD ERRORS OF MEANS, MEDIANS, AND SAMPLE SIZES FOR ELIGIBLE CHILDREN IN SHORT-, MEDIUM-, AND LONG-TERM CENTERS

Sex	Race	Age	Short				Medium				Long			
			Mean	σ_M(1)	Median	N	Mean	σ_M	Median	N	Mean	σ_M	Median	N
Male	White	3	140.00	NC(3)	135.00	3	166.00	NC	166.00	1	-	-	-	0
		4	150.15	NC	146.50	8	138.41	NC	**	37	145.90	NC	144.00	10
		5	139.33	NC	137.00	21	149.47	NC	**	49	148.38	NC	148.50	8
		6	152.81	NC	152.00	25	137.91	NC	**	32	-	-	-	0
	Non-White	3	126.56	NC	123.00	9	125.28	NC	114.00	7	145.82	NC	147.00	11
		4	138.48	NC	**(2)	52	135.50	NC	131.50	20	140.34	NC	**	54
		5	137.11	NC	**	34	143.23	NC	144.00	9	150.34	NC	154.00	29
		6	154.60	NC	164.00	5	147.00	NC	147.00	2	168.00	NC	168.00	1
Female	White	3	143.00	NC	137.00	4	124.75	NC	122.00	4	127.00	NC	127.00	2
		4	146.05	NC	148.50	18	142.26	NC	**	31	148.80	NC	142.00	5
		5	141.87	NC	145.00	24	140.54	NC	**	43	142.67	NC	157.00	6
		6	145.36	NC	145.00	25	138.83	NC	142.00	24	-	-	-	0
	Non-White	3	147.50	NC	152.00	6	134.14	NC	138.00	7	137.72	NC	144.00	7
		4	149.80	NC	**	51	141.87	NC	140.00	15	142.73	NC	**	41
		5	143.96	NC	144.50	26	145.36	NC	142.00	11	149.68	NC	**	37
		6	138.62	NC	147.00	13	137.67	NC	132.00	3	175.00	NC	175.00	1
	Subtotal, by Age	3	137.09	NC	137.50	22	130.58	NC	122.00	19	141.10	NC	144.00	20
		4	144.74	NC	**	129	139.50	NC	**	103	139.65	NC	NC	110
		5	140.34	NC	**	105	145.13	NC	**	112	149.26	NC	NC	80
		6	147.50	NC	**	68	138.56	NC	**	61	171.50	NC	171.50	2
	Total	All ages	143.07	2.50	**	324	140.87	1.53	**	295	187.45	1.95	NC	212

Notes: (1) Standard error of the mean, computed only for total, all ages.
(2) The median was not obtained when N was more than 30.
(3) Not computed.

EXHIBIT 31 - PSI SUBTEST MEAN(1) RAW SCORES FOR ALL CHILDREN, BY AGE AND RACE

		3		4		5		6		Total Possible Score
		White	Non-White	White	Non-White	White	Non-White	White	Non-White	
Short-Term	Subtest 1	12.43	13.11	13.71	15.33	18.00	16.72	17.82	19.72	26
	Subtest 2	5.14	5.56	7.75	8.34	12.00	10.23	10.65	12.61	26
	Subtest 3	4.43	4.72	6.96	6.86	10.37	7.98	10.94	10.72	19
	Subtest 4	9.14	7.83	11.57	10.68	15.09	12.65	16.00	14.28	19
	Total	32.00	30.67	39.22	42.54	55.22	47.50	55.50	57.28	90
Medium-Term	Subtest 1	12.50	11.79	16.58	14.74	18.56	16.46	19.98	22.00	26
	Subtest 2	6.13	4.89	8.76	8.61	10.84	9.29	14.57	10.50	26
	Subtest 3	4.50	5.16	7.91	6.30	9.98	8.71	12.56	10.00	19
	Subtest 4	8.50	6.95	12.34	10.15	14.23	12.36	15.76	14.25	19
	Total	31.80	28.19	45.89	38.96	53.03	45.07	61.89	59.53	90
Long-Term	Subtest 1	10.67	12.00	12.81	15.32	15.26	17.49	-	19.25	26
	Subtest 2	4.00	5.17	9.62	8.00	11.53	10.16	-	10.25	26
	Subtest 3	3.00	5.50	7.38	6.96	10.00	8.13	-	10.25	19
	Subtest 4	6.67	8.70	11.06	11.11	13.16	12.97	-	15.50	19
	Total	24.34	30.46	40.90	41.04	49.95	48.22	-	62.50	90

Note: (1) Unadjusted.

EXHIBIT 32 - PSI SUBTEST MEAN(1) RAW SCORES FOR ELIGIBLE CHILDREN, BY AGE AND RACE

		3		4		5		6		Total Possible Score
		White	Non-White	White	Non-White	White	Non-White	White	Non-White	
Short-Term	Subtest 1	12.43	13.11	13.71	15.33	18.00	16.72	17.82	19.72	26
	Subtest 2	5.14	5.56	7.75	8.34	12.00	10.23	10.65	12.61	26
	Subtest 3	4.43	4.72	6.96	6.86	10.37	7.98	10.94	10.72	19
	Subtest 4	9.14	7.83	11.57	10.68	15.09	12.65	16.00	14.28	19
	Total	32.00	30.67	39.22	42.54	55.22	47.50	55.50	57.28	90
Medium-Term	Subtest 1	12.57	11.50	16.42	14.88	18.61	17.00	19.96	22.00	26
	Subtest 2	5.86	4.63	8.83	8.82	10.87	9.55	14.50	10.50	26
	Subtest 3	4.86	4.69	7.74	5.91	10.01	8.75	12.60	10.00	19
	Subtest 4	8.86	6.81	12.18	10.29	14.26	12.95	15.71	14.25	19
	Total	33.00	28.07	45.45	38.86	52.10	46.25	62.35	55.75	90
Long-Term	Subtest 1	14.00	11.95	13.23	15.34	15.14	17.57	-	18.67	26
	Subtest 2	6.00	5.52	11.00	8.06	11.50	10.48	-	10.33	26
	Subtest 3	6.00	5.10	7.23	6.89	9.71	8.11	-	10.67	19
	Subtest 4	9.67	8.95	11.54	11.22	13.00	13.02	-	15.33	19
	Total	32.00	29.50	42.34	41.11	42.35	48.97	-	62.00	90

Note: (1) Unadjusted.

EXHIBIT 33 - PSI SUBTEST MEAN(1) RAW SCORES EXPRESSED AS PERCENT OF TOTAL POSSIBLE SCORES FOR ALL CHILDREN

		3		4		5		6	
		White	Non-White	White	Non-White	White	Non-White	White	Non-White
Short-Term	Subtest 1	47.8	50.4	52.7	59.0	69.2	64.3	68.5	75.8
	Subtest 2	19.8	21.4	30.0	32.1	46.2	39.3	41.0	49.0
	Subtest 3	23.3	24.8	36.6	36.1	54.6	42.0	57.6	56.4
	Subtest 4	48.1	41.2	60.9	56.2	79.4	66.6	84.2	75.2
	Total	35.5	34.1	43.6	47.3	61.4	52.8	61.7	63.6
Medium-Term	Subtest 1	49.1	45.3	63.8	56.7	71.4	63.3	76.8	84.6
	Subtest 2	23.6	18.8	33.7	33.1	41.7	35.7	56.0	40.4
	Subtest 3	23.7	27.2	41.6	33.2	52.5	45.8	66.1	52.6
	Subtest 4	44.7	36.6	64.9	53.4	74.9	65.1	82.9	75.0
	Total	35.3	31.3	51.0	43.3	58.9	50.1	68.8	66.1
Long-Term	Subtest 1	41.0	46.2	49.2	58.9	58.7	67.3	-	74.0
	Subtest 2	15.4	19.9	37.0	30.8	44.3	39.1	-	39.4
	Subtest 3	15.8	28.9	38.8	36.6	52.6	42.8	-	54.9
	Subtest 4	35.1	45.8	58.2	58.5	69.3	68.3	-	81.6
	Total	27.0	33.8	45.4	45.6	55.5	53.6	-	69.4

Note: (1) Unadjusted.

EXHIBIT 34 - PSI SUBTEST MEAN[(1)] RAW SCORES EXPRESSED AS PERCENT OF TOTAL POSSIBLE SCORES FOR ELIGIBLE CHILDREN

		3		4		5		6	
		White	Non-White	White	Non-White	White	Non-White	White	Non-White
Short-Term	Subtest 1	47.8	50.4	52.7	59.0	69.2	64.3	68.5	75.8
	Subtest 2	19.8	21.4	30.0	32.1	46.2	39.3	41.0	49.0
	Subtest 3	23.3	24.8	36.6	36.1	54.6	42.0	57.6	56.4
	Subtest 4	48.1	41.2	60.9	56.2	79.4	66.6	84.2	75.2
	Total	35.5	34.1	43.6	47.3	61.4	52.8	61.7	63.6
Medium-Term	Subtest 1	48.3	44.2	63.2	57.2	71.2	65.4	76.8	84.6
	Subtest 2	22.5	17.8	34.0	33.9	41.8	36.7	55.8	40.3
	Subtest 3	25.6	24.7	40.7	31.1	52.7	46.1	66.3	52.6
	Subtest 4	46.6	35.8	64.1	54.2	75.1	68.2	82.7	75.0
	Total	36.6	31.2	50.5	43.2	57.9	51.4	69.3	61.9
Long-Term	Subtest 1	53.8	46.0	50.9	59.0	58.2	67.6	-	71.8
	Subtest 2	23.1	21.2	42.3	31.0	44.2	40.3	-	39.7
	Subtest 3	31.6	26.9	38.1	36.3	51.1	42.7	-	56.2
	Subtest 4	50.9	47.1	60.7	59.1	68.4	68.5	-	80.7
	Total	35.6	32.8	47.0	45.7	47.1	54.4	-	68.9

Note: (1) Unadjusted.

EXHIBIT 35 - PERCENTAGE OF ELIGIBLE CHILDREN PASSING PSI ITEMS, BY AGE AND PROGRAM TYPE

PSI SUBTEST 1

Item No.	3			4			5			6		
	S	M	L	S	M	L	S	M	L	S	M	L
1.	100.0	87.0	95.8	79.7	93.4	87.9	72.8	94.5	87.0	82.4	80.7	100.0
2.	24.0	34.8	29.2	21.9	50.0	25.2	35.0	53.2	42.9	44.1	40.4	33.3
3.	48.0	39.1	62.5	69.5	62.3	64.5	75.7	81.7	81.8	75.0	71.9	33.3
4.	12.0	4.4	16.7	17.2	20.8	21.5	23.3	20.2	18.2	17.7	8.8	66.7
5.	96.0	87.0	95.8	98.4	94.3	96.3	96.1	94.5	93.5	94.1	98.3	100.0
6.	96.0	91.3	95.8	94.5	87.7	89.7	94.2	90.8	92.2	88.2	93.0	100.0
7.	80.0	52.2	54.2	79.7	79.3	74.8	85.4	79.8	80.5	77.9	89.5	100.0
8.	32.0	21.7	37.5	57.0	60.4	57.9	69.9	70.6	61.0	72.1	73.7	66.7
9.	96.0	95.7	66.7	92.2	94.3	96.3	98.1	96.3	97.4	97.1	100.0	100.0
10.	72.0	73.9	70.8	78.9	86.8	81.3	84.5	85.3	89.6	92.7	89.5	66.7
11.	36.0	47.8	54.2	64.8	63.2	69.2	67.0	78.0	88.3	75.0	86.0	100.0
12.	28.0	17.4	20.8	39.1	49.1	42.1	59.2	64.2	54.6	50.0	71.9	66.7
13.	96.0	95.7	95.8	96.9	94.3	97.2	98.1	97.3	94.8	98.5	100.0	100.0
14.	48.0	26.1	37.5	63.3	57.6	42.1	68.0	72.5	58.4	86.8	87.7	66.7
15.	44.0	47.8	50.0	61.7	56.6	68.2	72.8	69.7	74.0	61.8	87.7	100.0
16.	40.0	34.8	50.0	55.5	54.7	54.2	68.9	60.6	74.0	52.9	84.2	66.7
17.	96.0	65.2	75.0	85.9	79.3	85.1	91.3	91.7	92.2	91.2	96.5	100.0
18.	92.0	60.9	91.7	87.5	89.6	96.3	93.2	93.6	94.8	95.6	96.5	100.0
19.	20.0	30.4	16.7	47.7	50.0	45.8	63.1	68.8	53.3	73.5	82.5	66.7
20.	00.0	4.4	12.5	21.1	21.7	23.4	24.3	38.5	29.9	50.0	61.4	33.3
21.	32.0	30.4	20.8	35.9	40.6	27.1	45.6	59.6	35.1	60.3	59.6	33.3
22.	8.0	34.8	12.5	18.0	16.0	15.0	20.4	21.1	22.1	25.0	35.1	0.0
23.	4.0	17.4	8.3	18.0	28.3	24.3	47.6	44.0	31.2	58.8	70.2	66.7
24.	52.0	26.1	20.8	54.7	66.0	54.2	75.7	83.5	79.2	79.4	89.5	100.0
25.	4.0	8.7	4.2	14.8	23.6	19.6	30.1	43.1	27.3	54.4	61.4	33.3
26.	36.0	47.8	25.0	43.8	72.6	49.5	65.1	78.0	59.7	83.8	84.2	66.7
Number of Children	25	23	24	128	106	107	103	109	77	68	57	3

EXHIBIT 35 (Continued)

PSI SUBTEST 2

Item No.	3			4			5			6		
	S	M	L	S	M	L	S	M	L	S	M	L
27.	4.0	4.4	0.0	5.5	15.1	6.5	19.4	21.1	14.3	25.0	35.1	33.3
28.	0.0	4.4	0.0	5.5	12.3	5.6	13.6	22.0	10.4	17.7	45.6	0.0
29.	4.0	4.4	8.3	11.7	14.2	12.2	28.2	31.2	9.1	38.2	57.9	33.3
30.	0.0	8.7	16.7	18.0	17.0	30.8	22.3	22.0	37.7	14.7	29.8	66.7
31.	4.0	4.4	8.3	12.5	15.1	15.9	35.0	25.7	27.3	25.0	49.1	0.0
32.	28.0	13.0	20.8	26.6	29.3	29.0	61.2	40.4	50.7	50.0	79.0	66.7
33.	12.0	4.4	8.3	15.6	22.6	17.8	31.1	35.8	28.6	27.9	45.6	0.0
34.	44.0	30.4	25.0	49.2	40.6	48.6	53.4	63.3	62.3	51.5	70.2	33.3
35.	0.0	0.0	4.2	4.7	13.2	7.5	11.7	17.4	23.4	17.7	24.6	0.0
36.	4.0	0.0	0.0	3.9	10.4	5.6	8.7	13.8	9.1	14.7	22.8	33.3
37.	0.0	0.0	4.2	3.1	4.7	6.5	5.8	10.1	10.4	2.9	33.3	33.3
38.	16.0	21.7	3.3	36.7	37.7	50.5	58.3	53.2	64.9	61.8	77.2	66.7
39.	32.0	30.4	25.0	64.8	63.2	52.3	70.9	69.7	72.7	69.1	87.7	66.7
40.	56.0	43.5	54.2	74.2	62.3	56.1	71.8	70.6	62.3	76.5	87.7	100.0
41.	56.0	52.2	50.0	59.4	55.7	69.2	72.8	58.7	75.3	57.4	79.0	66.7
42.	32.0	52.2	41.7	57.0	62.3	55.1	77.7	64.2	70.1	60.3	82.5	100.0
2(1)	0.0	8.7	4.2	8.6	22.6	14.0	20.4	21.1	19.1	13.2	21.1	0.0
43. 1	16.0	17.4	16.7	35.9	35.8	41.1	38.8	48.6	58.8	45.6	54.4	66.7
T	16.0	26.1	20.9	44.5	58.4	55.1	59.2	69.7	77.9	58.8	75.5	66.7
2	0.0	4.3	0.0	8.6	11.3	5.6	9.7	7.3	19.1	17.6	15.8	0.0
44. 1	60.0	65.2	62.5	61.7	62.3	70.1	70.9	65.1	69.1	66.2	75.4	66.7
T	60.0	69.5	62.5	70.3	73.6	75.7	80.6	72.4	88.2	83.8	91.2	66.7
2	0.0	0.0	0.0	9.4	15.1	14.0	20.4	16.5	17.6	33.8	17.5	0.0
45. 1	4.0	26.1	41.7	44.5	43.4	37.4	51.5	48.6	47.1	47.1	61.4	33.3
T	4.0	26.1	41.7	53.9	58.5	51.4	71.9	65.1	64.7	80.9	78.9	33.3
2	8.0	0.0	0.0	14.8	7.5	6.5	19.4	9.2	11.8	41.2	14.0	0.0
46. 1	56.0	34.8	41.7	50.8	64.2	53.3	55.3	67.9	69.1	32.4	64.9	66.7
T	64.0	34.8	41.7	65.6	71.7	59.8	74.7	77.1	80.9	73.6	78.9	66.7
2	4.0	0.0	4.2	8.6	10.4	6.5	11.7	11.0	14.7	13.2	21.2	33.3
47. 1	56.0	56.5	75.0	78.9	73.6	77.6	75.7	82.6	86.8	77.9	77.2	33.3
T	60.0	56.5	79.2	87.5	84.0	84.1	87.4	93.6	101.5	91.1	98.4	66.6
Number of Children	25	23	24	128	106	107	103	109	77	68	57	3

Note: (1) For items 43 through 47, 2 represents the percentage of children getting 2 points, 1 represents the percentage of children getting 1 point, and T represents the percentage of children getting 1 or 2 points.

EXHIBIT 35 (Continued)

PSI SUBTEST 3

Item No.	3			4			5			6		
	S	M	L	S	M	L	S	M	L	S	M	L
48.	44.0	73.9	70.8	89.1	80.2	83.2	95.2	92.7	92.2	97.1	96.5	100.0
49.	60.0	69.6	66.7	70.3	70.8	69.2	75.7	84.4	71.4	88.2	89.5	66.7
50.	44.0	39.1	58.3	67.2	62.3	54.2	76.7	78.9	70.1	82.4	91.2	100.0
51.	0.0	8.7	0.0	3.1	1.9	1.9	7.8	6.4	3.9	25.0	7.0	33.3
52.	24.0	17.4	8.3	21.9	27.4	29.0	42.7	45.0	39.0	69.1	70.2	66.7
53.	36.0	21.7	37.5	60.9	57.6	49.5	74.8	71.6	57.1	79.4	80.7	100.0
54.	8.0	8.7	12.5	26.6	27.4	25.2	46.6	48.6	28.6	54.4	68.4	100.0
55.	24.0	0.0	20.8	26.6	27.4	16.8	29.1	41.3	9.1	26.5	54.4	0.0
56.	0.0	8.7	8.3	5.5	10.4	9.4	13.6	17.4	6.5	13.2	15.8	0.0
57.	68.0	43.5	50.0	71.1	66.0	77.6	83.5	80.7	92.2	94.1	91.2	100.0
58.	16.0	4.4	16.7	25.0	25.5	22.4	37.9	33.0	29.9	42.7	59.7	66.7
59.	64.0	69.6	66.7	73.4	81.1	74.8	86.4	89.9	80.5	86.8	86.0	100.0
60.	4.0	4.4	0.0	5.5	7.6	13.1	10.7	27.5	15.6	11.8	33.3	0.0
61.	28.0	56.5	37.5	34.4	45.3	32.7	38.8	44.0	42.9	52.9	54.4	66.7
62.	8.0	13.0	16.7	43.8	41.5	41.1	58.3	66.1	62.3	66.2	87.7	66.7
63.	20.0	13.0	33.3	34.4	34.9	44.9	49.5	63.3	54.6	70.6	86.0	66.7
64.	8.0	8.7	8.3	21.1	24.5	24.3	35.9	50.5	40.3	61.8	82.5	33.3
65.	0.0	8.7	0.0	4.7	15.1	9.4	16.5	21.1	22.1	44.1	38.6	0.0
66.	8.0	4.4	8.3	3.9	8.5	15.0	18.5	15.6	22.1	29.4	29.8	0.0
Number of Children	25	23	24	128	106	107	103	109	77	68	57	3

EXHIBIT 35 (Continued)

PSI SUBTEST 4

Item No.	3			4			5			6		
	S	M	L	S	M	L	S	M	L	S	M	L
67.	92.0	73.9	91.7	89.8	94.3	96.3	94.2	96.3	92.2	100.0	100.0	66.7
68.	72.0	56.5	70.8	82.0	90.6	93.5	95.2	96.3	93.5	97.1	100.0	100.0
69.	16.0	13.0	20.8	39.1	49.1	52.3	63.1	77.1	75.3	92.7	91.2	100.0
70.	4.0	0.0	8.3	23.4	26.4	30.8	42.7	52.3	49.4	73.5	77.2	66.7
71.	76.0	60.9	83.3	84.4	89.6	84.1	94.2	89.9	93.5	94.1	89.5	100.0
72.	48.0	34.8	66.7	53.9	56.6	53.3	65.1	69.7	59.7	77.9	66.7	100.0
73.	68.0	43.5	54.2	68.8	67.0	59.8	82.5	77.1	71.4	88.2	89.5	66.7
74.	40.0	60.9	41.7	50.0	60.4	57.9	68.0	62.4	61.0	82.4	82.5	33.3
75.	52.0	43.5	41.7	74.2	75.5	72.9	86.4	85.3	75.3	94.1	89.5	100.0
76.	20.0	39.1	45.8	50.8	50.0	43.9	54.4	59.6	63.6	55.9	64.9	33.3
77.	40.0	39.1	33.3	60.9	52.8	58.9	76.7	77.1	74.0	86.8	73.7	100.0
78.	56.0	47.8	54.2	62.5	49.1	56.1	66.0	63.3	54.6	67.7	70.2	100.0
79.	52.0	47.8	50.0	64.8	74.5	69.2	87.4	81.7	77.9	86.8	89.5	100.0
80.	48.0	47.8	62.5	64.1	66.0	67.3	84.5	84.4	77.9	85.3	96.5	100.0
81.	20.0	13.0	25.0	24.2	34.0	25.2	38.8	45.9	27.3	57.4	64.9	33.3
82.	20.0	30.4	54.2	47.7	57.6	53.3	68.0	67.9	61.0	70.6	68.4	100.0
83.	24.0	43.5	50.0	57.0	61.3	57.0	71.8	76.2	70.1	88.2	84.2	100.0
84.	20.0	13.0	8.3	36.7	37.7	36.5	55.3	56.0	48.1	69.1	73.7	66.7
85.	52.0	34.8	41.7	53.1	65.1	57.9	72.8	83.5	75.3	86.8	87.7	66.7
Number of Children	25	23	24	128	106	107	103	109	77	68	57	3

each age level and program type) who passed the item is listed. For items 43 through 47 in subtest 2, the percentage of children receiving 2 points or 1 point is shown separately, plus the percentage (T) of children receiving either 1 or 2 but not 0.

Comparisons within a column--that is, between items--are based on responses of the same sample of children. Comparisons across columns are comparisons between groups. The reader should thus take into account changes in denominators.

2. Statistical Analyses

This subsection will present the results of covariance and regression analyses performed on the data. While a large number of analyses and results are given, the reader will note that they are simply variations of two basic methods of analysis with two types of observation.[1] Thus, there are:

a. Analyses based on the CDC as the unit of observation:
   (1) Multivariate analysis of variance with multiple covariance adjustments.[2]
   (2) Multivariate multiple regression analysis.[3]

b. Analyses based on the individual child as the unit of observation:
   (1) Multivariate analysis of covariance.[4]
   (2) Univariate stepwise regression analysis.[5]

For convenience, the first will be referred to as "By-Center" analyses, and the second as "By-Child" analyses.

---

[1] See the detailed discussion of statistical models in Appendix B.

[2] Cooley, William W. and Paul R. Lohnes, Multivariate Procedures for the Behavioral Sciences (Chapter 4). New York, Wiley and Sons, Inc., 1962

[3] Anderson, T.W., Introduction to Multivariate Statistical Analysis (Section 8.2). New York: Wiley and Sons, Inc., 1958

[4] Cooley, loc. cit.

[5] Intertest correlations were eventually calculated for one of these analyses, thus making it in effect multivariate.

a. By-Center Analyses

(1) Analyses of Covariance with All Eligible Children Tested in English[1]

The first covariance analysis made used the following covariates and measures for each of 69 CDC's:

- Age: the average age in months of the center's sample of eligible children.
- Sex: the proportion of males in the center's sample.
- Race: the proportion of non-whites in the center's sample.
- Population (Community Size): the log of the population of the CDC's town.
- Poverty: the proportion of children defined as poor in the center's sample.

The classification of a child as "poor" or "not poor" was made by entering in Exhibit 36 the income and household size data from the Family Information Form filled out by parents.[2]

---

[1] The original total sample of eligibles contained 831 children from 71 CDC's. However, 67 of these children were tested in Spanish. In early covariance analyses, language of testing was included as a covariate. It was found, however, that because of the very badly skewed distribution of proportion of children tested in Spanish across S, M, and L centers, adjustments were being made in means that were disproportionately large compared with those for other covariates such as age and race. Therefore, the Spanish-speaking children were dropped from the sample for subsequent By-Center analyses. A By-Child stepwise regression analysis was made for this subgroup; the results of that analysis are reported in subsection IV.B.2.b.(2). (For a thorough discussion of the problem of selection and use of covariates see Cochran, W.G., "Analysis of Covariance: Its Nature and Uses," Biometrics, September 1957, pp. 261-281.) Unless otherwise indicated, all analyses reported are based on the children tested in English.

[2] The poverty index used here is not identical with OEO's index, since the income increments on CAP-HS Form 46 or 46a (dated 20 June 1966) are different from those required by the OEO poverty scale. In addition, no attempt was made to distinguish between farm and non-farm children or to make regional adjustments. If all the center's eligible children could not be classified because the Family Information Form was incomplete or missing, the proportion of classifiable eligible children was used.

EXHIBIT 36 - POVERTY AS A FUNCTION OF HOUSEHOLD SIZE AND INCOME

Household Size	Income				
	$1,000 to 1,999	$2,000 to 2,999	$3,000 to 3,999	$4,000 to 4,999	$5,000 or Over
1	Poor	NP(1)	NP	NP	NP
2	Poor	NP	NP	NP	NP
3	Poor	Poor	NP	NP	NP
4	Poor	Poor	NP	NP	NP
5	Poor	Poor	Poor	NP	NP
6	Poor	Poor	Poor	NP	NP
7	Poor	Poor	Poor	Poor	NP
8 or more	Poor	Poor	Poor	Poor	NP

Note: (1) NP denotes "not poor."

The dependent variable is a 4-variate observation of the form:

$$Y_{ij} = \begin{array}{l} \text{Center average PPVT raw score} \\ \text{Center average PSI total raw score} \\ \text{Center average BI total raw score} \\ \text{Center average VSMS raw score} \end{array}$$

where $j = S, M, L$
$i = 1, 2, 3, \ldots, n_j$
$n_S = 22$
$n_M = 26$
$n_L = 21$

An adjusted mean $\hat{\bar{Y}}_{jk}$ for measure k (e.g., PPVT) for condition j (S, M, or L) is obtained by:

$$\hat{\bar{Y}}_{jk} = \bar{Y}_{jk} - \left[b_1 (\bar{X}_{ij} - \bar{X}_{1.}) + b_2 (\bar{X}_{2j} - \bar{X}_{2.}) + \ldots + b_c (\bar{X}_{cj} - \bar{X}_{c.}) \right]$$

where $\overline{Y}_{jk}$ = the unweighted mean of measuring k in treatment level j

$\overline{X}_{cj}$ = the unweighted average of control variable X_c for treatment level j

$\overline{X}_{c.}$ = the total unweighted average of c over all conditions

b_c = the beta coefficient for control variable c

Exhibit 37 shows obtained and estimated parameters of the me sures or dependent variables for the total sample of 69 CDC's.

EXHIBIT 37 - TOTAL SAMPLE PARAMETERS FOR DEPENDENT VARIABLES

Dependent Variable	Total Sample Mean	Sample Standard Deviation	Estimate of Center Standard Deviation
PPVT	39.63	5.99	4.39
PSI	46.75	8.65	5.22
BI	143.33	8.95	8.80
VSMS	55.28	4.77	4.57

Exhibit 38 lists the total unweighted mean and standard deviation of each covariate. The reader should recall that these are based on the distribution of center parameters in each case.

EXHIBIT 38 - COVARIATE MEANS AND STANDARD DEVIATIONS

Covariate	Total Sample Mean	Standard Deviation
Age in months	60.00	6.49
Percentage of males	51.90	16.78
Percentage of non-white children	56.73	33.81
Log of population of town	5.27	1.04
Percentage of poor	39.05	25.21

Exhibit 39 shows the unadjusted and adjusted means for each test for S, M, and L centers. It also shows the total increment added to or subtracted from each unadjusted mean, along with the components of the total increment provided by each control variable or covariate. For example, the unweighted average PPVT center average raw score for the S centers was 38.64. The adjusted mean was 39.03, obtained by addition of the algebraic sum of the components to the unweighted mean (the component signs shown are the resultants of $-\left[b_c(\bar{X}_{cj} - \bar{X}_{.c})\right]$). Thus, 38.64 plus 0.39 equals 39.03; 41.73 minus 2.16 equals 39.57; etc.

The test criterion for the multivariate covariance analysis is the Wilks' Lambda, Λ, which is like the F-ratio of a univariate analysis of variance. In this analysis, Λ equals 0.8511. A transformation of this yields:

$$F_{8,116} = 1.23\ ,\ p > 0.05$$

Thus, the analysis does not support the hypothesis that there are significant differences between the means of the S, M, and L centers for the 4-variate dependent variable.[1]

The F-ratio for each measure taken individually was calculated. These were:

Test	$F_{2,61}$	Significance
PPVT	0.38	$p > 0.05$
PSI	0.04	$p > 0.05$
BI	3.47	$p < 0.05$
VSMS	1.20	$p > 0.05$

---

[1] The overall sample was designed to detect differences between PPVT raw score means of six to seven points with a power of about 0.70. In fact, the analysis permitted sufficient reduction of error variance to detect differences of five points with a power of about 0.80. It may be noted here also that there was no significant difference in statistical outcome when the same type of analysis included Spanish-tested eligibles and included language as a covariate.

EXHIBIT 39 - COVARIANCE ADJUSTMENTS TO TEST RAW SCORE MEANS BY COVARIATE AND DURATION LEVEL (S, M, L)

Test	Program Duration	Unweighted Mean	Adjusted Mean	Total Increment	= Age Increment	+ Sex Increment	+ Race Increment	+ Population Increment	+ Poverty Increment
PPVT	Short	38.64	39.03	0.39	0.0274	-0.0361	0.3084	-0.0365	0.1257
	Medium	41.73	39.57	-2.16	-1.0138	0.0121	-1.0886	-0.1459	0.0731
	Long	38.07	40.34	2.27	1.2275	0.0229	1.0247	0.2165	-0.2222
PSI	Short	46.50	46.48	-0.02	0.0494	-0.2064	0.1679	-0.0990	0.0780
	Medium	49.49	46.80	-2.69	-1.8267	0.0690	-0.5926	0.3958	0.0454
	Long	43.61	46.96	3.35	2.2118	0.1308	0.5578	0.5871	-0.1380
VSMS	Short	55.49	55.58	0.09	0.0168	-0.0752	-0.0348	0.0416	0.1431
	Medium	54.16	53.94	-0.22	-0.6226	0.0252	0.1228	0.1665	0.0833
	Long	56.44	56.63	0.19	0.7538	0.0477	-0.1156	-0.2470	-0.2529
BI	Short	144.22	145.34	1.12	0.0110	0.1265	0.6928	0.4380	0.1519
	Medium	139.98	138.74	-1.24	-0.4082	-0.0423	-2.4455	1.7520	-0.0878
	Long	146.54	146.90	0.36	0.4942	-0.0801	2.3019	-2.5988	0.2667

Thus, there was a significant difference in means for the Behavior Inventory, by the criterion of $\alpha = 0.05$. Examination of the adjusted means of the BI (see Exhibit 39 above) suggests that the mean of the M centers was significantly different from the means of S and L sample centers. The t-values for these two comparisons are significant at the 5 percent level (the t-value of the S-L contrast, of course, is not significant at that level). While the criterion $\alpha = 0.05$ is probably too liberal for the F-ratios, since we are making multiple, correlated comparisons, it is clear that the 5 to 6 percent difference from upper to lower BI means is unique among the four measures.

It is apparent in Exhibit 39 that for the individual tests (PPVT and PSI), the variables that generally contribute most to the adjustment of the means are race, population, and age. Poverty appears more related to PPVT than to PSI scores. With the Vineland Social Maturity Scale, poverty and population appear to take on more weight, and with the Behavior Inventory, the population and race components seem predominant.

(2) Multivariate Multiple Regression Analysis

To examine the relative contribution of different variables to the prediction of center average scores, PRC performed a multivariate multiple regression analysis. In this analysis, the treatment variable was included as a continuous variable, determined for a given center by dividing the sum of the number of weeks of enrollment in the program of all children in the sample by the number of children in the sample.

The model for this analysis was of the form:

$$\hat{\bar{Y}} = \alpha + \beta_1 \bar{X}_1 + \beta_2 \bar{X}_2 + \ldots + \beta_c \bar{X}_c$$

The independent variables $(X_1 \ldots X_c)$ were the same as those used in the By-Center covariance analysis, plus the exposure variable defined above and measured in weeks.

Exhibit 40 shows, for each measure, the estimates of the constant term α, the β coefficients, and the standard deviation of the coefficient. The unweighted center average for each test is listed for convenience.

A question of immediate interest is whether or not the β coefficient for a given variable is really different from zero. To measure this probability, the 95 percent confidence limits for each β were calculated.[1] Exhibit 41 shows for each measure those independent variables whose weight (β) had a confidence interval that did not include zero.

The percentage of total variance for each test attributed to the six independent variables taken together is approximately:

Test	Percent
PPVT	62
PSI	74
BI	9
VSMS	14

The PPVT was used to make some checks on the adequacy of the model; in particular, examinations of a possible non-linear effect of age and the general size of a sex-race interaction were desirable. These factors were investigated using the By-Center data calculated from the total sample; including the children tested in Spanish.[2] A calculation was first made of the percentage of total variance of the

---

[1] The reader can readily calculate this interval for himself from the parameters given in Exhibit 40.

[2] In many respects, parameters of the two samples (Total Sample of Children Tested in English and Total Sample of Children Tested) were very similar. Means and variances of the independent variables other than language of testing were similar. Wilk's Lambda for the covariance analysis was about the same and not significant in either case. The total percentage of dependent variable variance accounted for by the independent variance was much the same in both cases. For these reasons, it seems reasonably safe to conclude that conclusions about a sex-race interaction and a non-linear age effect drawn from one sample probably apply to the other.

EXHIBIT 40 - MULTIPLE REGRESSION PARAMETERS FOR EACH TEST

Test	Independent Variable	Estimate of β Coefficient	σ_β	α (Constant Term[1] for Test Indicated)	Unweighted Average of Center Means for Test Indicated
PPVT	Age	0.5382	0.0920	10.96	39.61
	Sex	0.0178	0.0324		
	Race	-0.0460	0.0155		
	Population	-0.2711	0.5676		
	Poverty	-0.0377	0.0233		
	Exposure	0.0560	0.0777		
PSI	Age	0.9584	0.1090	-3.55	46.84
	Sex	-0.0251	0.0384		
	Race	-0.0324	0.0183		
	Population	-0.8095	0.6723		
	Poverty	-0.0281	0.0276		
	Exposure	0.0762	0.0921		
BI	Age	0.2850	0.2025	113.02	142.86
	Sex	0.0131	0.0714		
	Race	-0.0145	0.0340		
	Population	2.7182	1.2496		
	Poverty	0.0228	0.0512		
	Exposure	-0.1142	0.1711		
VSMS	Age	0.3133	0.1071	36.69	55.21
	Sex	-0.0238	0.0377		
	Race	0.0184	0.0180		
	Population	0.0332	0.6606		
	Poverty	-0.0288	0.0271		
	Exposure	0.0466	0.0905		

Note: (1) Y-intercept of regression line.

EXHIBIT 41 - INDEPENDENT VARIABLES WITH SIGNIFICANT BETA COEFFICIENTS, BY TEST

Dependent Variable	Independent Variable					
	Age	Sex	Race	Exposure	Population	Poverty
PPVT	X(1)		X			
PSI	X					
BI					X	
VSMS	X					

Note: (1) An X indicates that the 95 percent confidence interval for the independent variable's β coefficient does not include zero.

center average PPVT raw scores attributable to Age (A), Sex (S), and Race (R) combined. The exposure variable W, the non-linear age variable A^2, and finally a sex-race interaction component SR were then added. The percentages of total variance attributable to these variables were approximately:

A, S, R	54.6
A, S, R, W	56.6
A, S, R, W, A^2	57.9
A, S, R, W, A^2, SR	58.2

In effect, by this procedure, the addition of the exposure variable increased the variance accounted for by about 2 percent of the total variance. The non-linear age variable accounted for another 1.3 percent, and the sex-race interaction component for yet another 0.3 percent. In these terms, it appears that the exposure variable adds little to center average prediction, and that prediction for the PPVT is not substantially influenced by non-linear age effects or by sex-race interactions.

(3) Multivariate Multiple Regression Analysis With Total Estimated Exposure as an Independent Variable

The exposure variable in the first regression analysis was based on weeks of enrollment of a child at time of testing. The daily length of programs varied to some extent (from 2 to 7 hours). To examine the possibility that total hours of exposure was a more significant measure than number of weeks regardless of hours per week, the exposure variable was converted and a By-Center multivariate multiple regression analysis was performed. This analysis utilized the total sample of children, including those tested in Spanish. The analysis also used language of testing as an independent variable.

In this analysis, the exposure variable was obtained by multiplying the center average weeks for the total center sample by five times the number of daily hours of the center's program. No attempt was made to determine the actual hours of attendance of each child. Thus, the measure is an approximation of the actual amount of exposure.

The results of this analysis were essentially the same as those of the primary regression analysis, described in subsection IV. B. 2. a. (2) above. As before, the 95 percent confidence limits for the β coefficient of the exposure variable included zero. Thus, no support was found for the hypothesis that total amount of exposure was any more significant a predictor than program length.

b. By-Child Analyses

(1) Analyses of Covariance With Eligible English-Tested Children in Subpopulations Defined by Age, Sex, and Race

Four-variate analyses of covariance, similar to the analyses described in subsections IV. B. 2. a. (1) and IV. B. 2. a. (2), were made with 10 subgroups of children defined by age, sex, and race. For these analyses, the observations were individual scores, rather than center averages. In performance of these analyses, four covariates were used:

- Family income
- Family size
- Population-community size (again expressed as the log of the population of the child's town)
- Daily program (number of hours per day of the child's CDC program)

Since data for the two family covariates came from the Family Information Form collected by the Bureau of the Census by mail, there were losses of children from the subsamples owing to missing data. In some cases, forms were not returned by parents. In some cases, incomplete forms were returned. Since, for the By-Child analyses, a child was dropped from the analysis if there were any missing data, the resultant sample sizes were substantially reduced. No analysis of the types of children for whom family data were missing has been made. The 10 groups analyzed, with the number of children in each group, are listed in Exhibit 42.

For convenience of reference, data on the independent variables and results for each group have been assembled in a single exhibit. Thus, Exhibit 43 shows, by group and treatment level within each group, the average for each covariate, and the unweighted and adjusted test means.

A group is starred if the Wilk's Lambda was significant at the 0.05 level. None of the F-ratios for the individual dependent variables was significant at the 5 percent level.

Finally, the total and component covariate increments for each group, level, and test are listed in Exhibit 44.

(2) Multiple Stepwise Regression Analyses

Multiple stepwise regression analyses were made with a number of dependent and independent variables. These analyses, unlike the By-Child covariance analyses, were univariate. The dependent variables analyzed were:

- PPVT Raw Score
- PSI Total Raw Score
- BI Total Raw Score
- VSMS Raw Score
- PSI Subtest 1 (Personal-Social Responsiveness)
- PSI Subtest 2 (Associative Vocabulary)
- PSI Subtest 3 (Numerical Concept Activation)
- PSI Subtest 4 (Sensory Concept Activation)
- PPVT IQ
- VSMS Social Quotient

It is convenient to think of the independent variables in terms of groups or subsamples of children as well as of names or labels. The groups, identified by name and number, are listed below with the independent variables (and their identification number) used with each group.

EXHIBIT 42 - SUBGROUPS USED FOR BY-CHILD COVARIANCE ANALYSES

Group	Number of Children in Sample			
	Short	Medium	Long	Total
1. 3-year-old males, non-white	6	4	8	18
2. 3-year-old females, non-white	2	5	5	12
3. 4-year-old males, white	8	22	3	33
4. 4-year-old males, non-white	37	10	40	87
5. 4-year-old females, white	12	18	2	32
6. 4-year-old females, non-white	33	9	26	68
7. 5-year-old males, white	9	35	6	50
8. 5-year-old males, non-white	21	6	15	42
9. 5-year-old females; white	18	33	4	55
10. 5-year-old females, non-white	21	7	23	51

EXHIBIT 43 - INDEPENDENT VARIABLE AND UNADJUSTED AND ADJUSTED DEPENDENT VARIABLE MEANS BY GROUP AND PROGRAM DURATION

Group	Duration	Average Log Population	Average Family Size	Average Family Income (dollars)	Average Hours Per Day	Unadjusted Average PPVT	Adjusted Average PPVT	Unadjusted Average PSI	Adjusted Average PSI	Unadjusted Average BI	Adjusted Average BI	Unadjusted Average VSMS	Adjusted Average VSMS
1. 3-Year-Old Males, Non-White	S	5.55	7.17	3,920	3.08	25.17	24.28	26.50	33.15	128.67	132.95	43.42	50.45
	M	5.65	5.75	4,500	5.25	18.50	16.34	22.50	16.72	129.75	120.88	48.50	46.50
	L	6.30	5.00	3,060	6.12	28.37	30.12	27.25	25.15	137.62	138.85	47.06	42.78
2. 3-Year-Old Females, Non-White*(1)	S	5.63	8.50	3,500	3.25	16.00	-8.78	27.50	24.97	174.50	183.76	50.00	56.72
	M	4.80	5.25	4,620	3.75	21.00	19.87	31.25	28.50	128.25	123.30	51.87	47.50
	L	6.40	5.60	4,400	4.70	28.00	38.82	29.40	32.61	136.80	137.06	53.90	54.71
3. 4-Year-Old Males, White	S	4.79	6.25	3,060	3.69	35.13	33.80	44.87	43.40	151.25	151.07	55.75	56.88
	M	4.70	7.50	4,450	3.68	43.27	43.83	45.68	45.25	136.45	135.64	53.14	52.76
	L	5.79	6.50	6,380	5.25	43.00	42.60	46.75	52.06	138.50	143.32	56.00	55.82
4. 4-Year-Old Males, Non-White	S	5.56	7.08	4,070	3.32	36.16	36.63	41.08	41.26	136.84	139.73	53.76	54.26
	M	5.18	6.70	3,100	4.50	36.30	35.45	40.50	40.60	140.00	138.63	51.60	53.36
	L	6.21	5.85	3,990	4.36	36.67	36.46	41.17	40.98	140.40	138.06	54.47	53.58
5. 4-Year-Old Females, White	S	5.04	7.08	3,830	3.83	35.42	34.02	41.25	41.17	142.08	145.46	56.83	56.72
	M	4.69	5.50	2,720	3.75	42.11	43.24	48.89	48.69	141.06	137.70	54.33	54.40
	L	5.27	8.50	5,000	4.50	38.50	36.74	51.00	53.31	139.50	149.45	54.00	54.09
6. 4-Year-Old Females, Non-White	S	5.71	5.67	4,450	3.44	36.85	36.14	43.48	42.74	148.55	147.99	53.88	53.40
	M	5.20	5.89	2,250	4.00	33.11	34.85	40.33	40.55	137.11	141.06	50.44	52.00
	L	6.22	6.85	3,620	3.54	37.69	37.99	42.81	43.68	145.46	144.80	56.08	56.29
7. 5-Year-Old Males, White	S	3.95	6.44	3,720	4.17	44.22	43.94	63.56	61.86	141.11	140.97	59.39	60.07
	M	4.48	6.77	3,630	3.60	46.66	46.96	52.34	52.22	145.97	146.95	54,19	53.92
	L	5.65	8.33	5,170	4.25	49.67	48.34	52.43	56.09	153.83	148,34	56.50	57.04
8. 5-Year-Old Males, Non-White	S	5.43	7.38	3,890	3.19	36.43	35.21	47.14	46.97	137.05	135.48	58.51	58.35
	M	4.81	5.50	2,420	3.92	33.67	34.55	37.83	37.47	133.67	128.24	52.50	53.20
	L	6.33	7.33	4,750	3.57	42.73	44.09	47.53	47.93	151.33	155.70	57.27	57.22
9. 5-Year-Old Females, White	S	4.77	6.11	3,720	3.61	42.00	41.97	55.50	55.56	149.89	148.86	58.25	58.77
	M	4.47	8.24	3,700	3.48	46.94	46.66	56.67	56.27	146.09	146.66	55.04	54.58
	L	5.34	7.50	5,630	3.13	45.25	47.69	56.25	59.30	158.00	157.91	63.87	65.32
10. 5-Year-Old Females, Non-White	S	5.30	7.19	3,600	3.45	39.62	38.85	49.05	48.87	145.00	145.55	57.79	57.05
	M	5.26	5.00	4,210	4.43	40.00	36.48	54.29	50.74	149.71	148.43	51.73	51.83
	L	6.10	5.96	3,780	3.61	38.09	39.86	48.26	49.50	154.87	154.76	58.65	59.30

Note: (1) An asterisk denotes that Wilk's Lambda is significant at 0.05 level.

EXHIBIT 44 - SUBGROUP COVARIANCE INCREMENTS

GROUP 1: THREE-YEAR-OLD MALES, NON-WHITE

Test	Duration	Population	Income	Family Size	Hours	Total
PPVT	S	-1.8391	-0.2882	4.3651	-3.1530	-0.85
	M	-1.3283	-0.9568	-0.4774	0.5655	-2.16
	L	1.9924	0.7032	-3.0351	2.0563	1.75
PSI	S	-4.1135	-0.3272	3.0889	7.9930	6.65
	M	-2.9709	-1.0863	-0.3379	-1.4335	-5.78
	L	4.4563	0.7984	-2.1478	-5.2128	-2.10
BI	S	-2.7713	-1.7718	8.0963	0.7188	4.28
	M	-2.0015	-5.8824	-0.8855	-0.1289	-8.87
	L	3.0023	4.3232	-5.6295	-0.4688	1.23
VSMS	S	-0.5927	-0.0986	2.1199	5.6239	7.03
	M	-0.4281	-0.3273	-0.2319	-1.0086	-2.00
	L	0.6421	0.2406	-1.4740	-3.6678	-4.28

GROUP 2: THREE-YEAR-OLD FEMALES, NON-WHITE

Test	Duration	Population	Income	Family Size	Hours	Total
PPVT	S	-0.4810	-2.0616	-22.0441	-0.1757	-24.78
	M	-8.4650	0.7542	6.6132	-0.0711	-1.13
	L	6.9259	0.2011	3.5271	0.1276	10.82
PSI	S	-0.2003	0.0134	-2.4304	0.1012	-2.53
	M	-3.5261	-0.0049	0.7291	0.0410	-2.75
	L	2.8850	-0.0013	0.3889	-0.0735	3.21
BI	S	-0.1256	-1.6859	11.1527	-0.0758	9.26
	M	-2.2097	0.6168	-3.3458	-0.0307	-4.95
	L	1.8080	0.1645	-1.7844	0.0551	0.26
VSMS	S	-0.1345	0.8533	5.8591	0.1501	6.72
	M	-2.3677	-0.3122	-1.7577	0.0608	-4.37
	L	1.9372	-0.0833	-0.9375	-0.1090	0.81

EXHIBIT 44 (Continued)

GROUP 3: FOUR-YEAR-OLD MALES, WHITE

Test	Duration	Population	Income	Family Size	Hours	Total
PPVT	S	0.0511	0.3298	-1.4574	-0.2525	-1.33
	M	0.1277	-0.0256	0.7114	-0.2666	0.56
	L	-0.8002	-0.5189	-1.0237	1.9361	-0.40
PSI	S	-0.1502	-0.0110	-0.8624	-0.4617	-1.47
	M	-0.3756	0.0009	0.4209	-0.4874	-0.43
	L	2.3536	0.0173	-0.6057	3.5398	5.31
BI	S	0.1509	0.8466	-0.0581	-1.1141	-0.18
	M	0.3773	-0.0656	0.0284	-1.1760	-0.81
	L	-2.3645	-1.3322	-0.0408	8.5416	4.82
VSMS	S	-0.0491	1.3848	-0.0462	-0.1628	1.13
	M	-0.1227	-0.1074	0.0225	-0.1718	0.38
	L	0.7686	-2.1791	-0.0324	1.2481	0.18

GROUP 4: FOUR-YEAR-OLD MALES, NON-WHITE

Test	Duration	Population	Income	Family Size	Hours	Total
PPVT	S	-0.3794	-0.0683	0.4531	0.4794	0.47
	M	-0.9561	0.3732	0.1708	-0.4330	-0.85
	L	0.6070	-0.0319	0.4605	-0.3248	-0.21
PSI	S	0.0069	-0.0039	0.1723	0.0075	0.18
	M	0.0173	0.0211	0.0650	-0.0067	0.10
	L	-0.0110	-0.0018	-0.1751	-0.0050	-0.19
BI	S	0.1537	-0.0268	0.4943	2.2978	2.89
	M	0.3873	0.1467	0.1864	-2.0754	-1.37
	L	-0.2459	-0.0125	-0.5024	-1.5565	-2.34
VSMS	S	0.4927	-0.0715	0.1388	-0.0813	0.50
	M	1.2416	0.3907	0.0523	0.0735	1.76
	L	-0.7883	-0.0334	-0.1411	0.0551	-0.89

EXHIBIT 44 (Continued)

GROUP 5: FOUR-YEAR-OLD FEMALES, WHITE

Test	Duration	Population	Income	Family Size	Hours	Total
PPVT	S	0.5650	0.1495	-2.1473	0.0000	-1.40
	M	-0.5336	-0.1522	2.0936	-0.2871	1.13
	L	1.2869	0.4676	-5.9587	2.4043	-1.76
PSI	S	0.3868	0.1869	-0.6879	0.0000	-0.08
	M	-0.3653	-0.1903	0.6707	-0.3251	-0.20
	L	0.8811	0.5845	-1.9088	2.7223	2.31
BI	S	-0.7302	2.4734	1.6486	0.0000	3.38
	M	0.6897	-2.5186	-1.6074	0.0781	-3.36
	L	1.6633	7.7356	4.5749	-0.6544	9.95
VSMS	S	-0.1518	0.1173	-0.0756	0.0000	-0.11
	M	0.1433	-0.1195	0.0737	-0.0346	0.07
	L	-0.3457	0.3669	-0.2098	0.2896	0.09

GROUP 6: FOUR-YEAR-OLD FEMALES, NON-WHITE

Test	Duration	Population	Income	Family Size	Hours	Total
PPVT	S	0.1046	-0.3898	-0.3053	-0.1153	-0.71
	M	0.5150	0.9208	-0.1654	0.4718	1.74
	L	-0.3058	0.1680	0.4453	-0.0105	0.30
PSI	S	-0.2290	-0.4061	-0.0196	-0.0930	-0.74
	M	-1.1274	0.9593	-0.0106	0.3806	0.22
	L	0.6694	0.1751	0.0286	-0.0085	0.87
BI	S	0.1748	-0.6564	0.2761	-0.3433	-0.56
	M	0.8603	1.5504	0.1496	1.4045	3.95
	L	-0.5108	0.2829	-0.4026	-0.0312	-0.66
VSMS	S	0.0154	-0.4977	-0.0214	0.0213	-0.48
	M	0.0757	1.1756	-0.0116	-0.0873	2.84
	L	-0.0449	0.2145	0.0312	0.0019	0.21

EXHIBIT 44 (Continued)

GROUP 7: FIVE-YEAR-OLD MALES, WHITE

Test	Duration	Population	Income	Family Size	Hours	Total
PPVT	S	0.0357	0.1724	-0.3312	-0.1643	-0.28
	M	0.0031	0.3135	-0.0936	0.0758	0.30
	L	-0.0689	-2.1006	1.0295	-0.1980	-1.33
PSI	S	-1.3850	-0.0011	-0.2173	-0.1085	-1.70
	M	-0.1194	-0.0019	-0.0614	0.0501	-0.12
	L	2.6746	0.0129	0.6755	-0.1308	3.26
BI	S	1.2533	0.3452	-0.7496	-0.9925	-0.14
	M	0.1080	0.6277	-0.2118	0.4581	0.98
	L	-2.4202	-4.2057	2.3302	-1.1961	-5.49
VSMS	S	0.0759	-0.0237	0.0843	0.5479	0.68
	M	0.0065	-0.0431	0.0238	-0.2529	-0.27
	L	-0.1465	0.2884	-0.2621	0.6603	0.54

GROUP 8: FIVE-YEAR-OLD MALES, NON-WHITE

Test	Duration	Population	Income	Family Size	Hours	Total
PPVT	S	-0.5460	0.0217	-0.3218	-0.3570	-1.22
	M	-2.0178	0.3407	1.8391	0.7288	0.88
	L	1.5905	-0.1649	-0.2644	0.2082	1.36
PSI	S	-0.3229	0.0459	-0.0923	0.2055	-0.17
	M	-1.1933	0.7200	0.5273	-0.4195	-0.36
	L	0.9406	-0.3485	-0.0758	-0.1199	0.40
BI	S	-1.4692	0.0644	0.3502	-0.4898	-1.57
	M	-5.4297	1.0109	-2.0009	1.0000	-5.43
	L	4.2799	-0.4894	0.2876	0.2857	4.37
VSMS	S	0.08147	0.0227	0.0878	-0.3504	-0.16
	M	0.3011	0.3563	-0.5014	0.7154	0.70
	L	-0.2373	-0.1725	0.0721	0.2044	-0.05

EXHIBIT 44 (Continued)

GROUP 9: FIVE-YEAR-OLD FEMALES, WHITE

Test	Duration	Population	Income	Family Size	Hours	Total
PPVT	S	0.1400	-0.1095	0.0049	-0.0679	-0.03
	M	-0.1600	-0.1264	-0.0027	0.0124	-0.28
	L	0.7098	1.4998	0.0000	0.2284	2.44
PSI	S	0.1836	-0.1210	0.1303	-0.1381	0.06
	M	-0.2098	-0.1397	-0.0708	0.0251	-0.40
	L	0.9309	1.6573	-0.0009	0.4647	3.05
BI	S	-0.3385	-0.1427	-0.6641	0.0993	-1.03
	M	0.3868	-0.1647	0.3609	-0.0181	0.57
	L	-1.7165	1.9539	0.0048	-0.3341	-0.09
VSMS	S	0.1986	-0.0196	0.3963	-0.0508	0.52
	M	-0.2270	-0.0227	-0.2154	0.0092	-0.46
	L	1.0072	0.2689	-0.0029	0.1708	1.45

GROUP 10: FIVE-YEAR-OLD FEMALES, NON-WHITE

Test	Duration	Population	Income	Family Size	Hours	Total
PPVT	S	-1.6516	0.1102	0.7430	0.0521	-0.77
	M	-1.8404	-0.3098	-1.1491	-0.1910	-3.52
	L	2.1235	-0.0138	-0.3197	0.0124	1.77
PSI	S	-1.4040	0.2162	1.1196	-0.1054	-0.18
	M	-1.5645	-0.6080	-1.7315	0.3865	-3.55
	L	1.8052	-0.0270	-0.4817	-0.0251	1.24
BI	S	-0.3595	0.2777	1.0182	-0.4029	0.55
	M	-0.4005	-0.7810	-1.5746	1.4773	-1.28
	L	0.4622	-0.0347	-0.4381	-0.0959	-0.11
VSMS	S	-0.5836	0.0128	0.0832	-0.2516	-0.74
	M	-0.6503	-0.0359	-0.1287	0.9225	0.10
	L	0.7504	-0.0016	-0.0358	-0.0599	0.35

- Group 1 utilized all eligible children tested in English (N = 517);
- Group 4 utilized all non-eligible children tested in English (N = 87); and
- Group 5 utilized children tested in Spanish (N = 42),

for these independent variables:

0 - Program Length in Weeks (Weeks)
1 - Age
2 - Mother's Education (M. Ed.)[1]
3 - Race
4 - Sex
5 - Daily Program Length in Hours (Hours)
6 - Population (Pop.)
7 - Family Size (F. Size)
8 - Center Size--Number of Classes in CDC--(C. Size)
9 - Family Income (Inc.)

- Group 2, obtained from Group 1, (N = 387) utilized variables 0 through 9, plus:

  10 - Father's Education (F. Ed.)[1]

- Group 3, obtained from Group 2, (N = 280), utilized variables 0 through 10, plus:

  11 - Child's CDC's teachers' average amount of experience with children from conditions of poverty (P-Exp.)[2]
  12 - Child's CDC's teachers' average amount of experience with pre-schoolers (PS-Exp.)[2]

---

[1] The variables for mother's education and father's education were simply numerical values for the scale provided on the Family Information Form:

1 = No school
2 = 1 to 3 years
3 = 4 to 6 years
4 = 7 to 8 years
5 = 9 to 11 years
6 = High school graduate
7 = Some college or college graduate.

[2] The variables for teachers' experience were simply numerical values given to the scales on the Staff Member Information Form:

1 = None
2 = 1 to 3 years
3 = 4 to 5 years
4 = Over 5 years

Not all dependent variables were analyzed for all groups. Highlights of the results of the analyses performed are given in Exhibit 45. Since Exhibit 45 summarizes a large amount of information quite compactly, a detailed explanation is warranted at this point.

Each row presents the outcomes of a given analysis. The first column on the left contains the name of the dependent variable used in the analysis (e.g., PPVT RS denotes PPVT Raw Score). The next column gives the group identification number and is followed by a column giving the group sample size. For the convenience of the reader, the average age (in months) of the group is presented next. The next two columns give the mean and standard deviations of the dependent variable for the group indicated. In the column following, the percentage of total score variance attributable to Variable 0 (program length in weeks) is given. "Weeks" was always forced into the analysis; consequently it always entered first. The next column (Total Percent of Variance Accounted for by All Independent Variables Used) gives, for the indicated analysis, $100 \times R^2$ from the final step. For example, for Group 1, all 10 independent variables together accounted about 32 percent of the PPVT Raw Score variance.

There follows a block of three columns containing t-values for the β coefficient for the "weeks" variable (Variable 0). The first t-value is that obtained initially, when the variable was forced into the analysis; the second value is the maximum in the analysis; the third is the final value at the end of the analysis. The sign, of course, corresponds to the sign of the β weight. For all practical purposes, in this exhibit an absolute value of 2.0 or more is associated with a probability of 0.95 or more that the coefficient is different from zero.

Exhibit 45 next lists the independent variables by number in the order in which they entered the analysis. It should be noted that the order starts with 2. Number 1 was omitted since, as noted above, it was always Variable 0, or Weeks. In a given analysis, the order in which each variable, other than Variable 0, entered the analysis depended in effect on which variable had the beta coefficient with the

EXHIBIT 45 - HIGHLIGHTS OF RESULTS OF STEPWISE REGRESSION ANALYSIS BY T

Dependent Variable	Group Number	Group N	Average Age	Dependent Variable Mean	Dependent Variable Standard Deviation	Initial % of Total Variance for Weeks	Total % of Variance for All Independent Variables Used	T-Value for β Coefficient for Weeks			Order of Entry of Independent Variables(2)									
								Starting	Maximum	Final	2	3	4	5	6	7	8	9	10	11
PPVT RS	1	517	60.18	40.83	11.28	0.10	31.73	-0.7075	0.7960	0.7960	8	1	21	22	17	23	2	19	18	-
	2	387	60.71	41.35	11.45	0.29	34.49	-1.0610	-1.0610	0.2878	8	1	20	21	17	19	23	2	18	22
	3	280	60.65	41.80	11.32	0.26	31.88	-0.8550	-1.3817	1.3574	8	20	1	21	15	16	17	2	19	18
	4	87	57.72	39.60	11.47	3.18	54.91	-1.6728	-1.6728	0.4734	19	8	22	23	21	1	17	18	2	-
	5	42	72.67	41.83	9.46	16.77	55.15	-2.8390	-2.8390	-2.2613	8	2	19	17	23	22	18	*(1)	-	-
PSI Total	1	517	60.18	47.83	14.92	0.23	41.68	-1.0982	1.3908	1.3572	8	21	1	19	2	18	22	17	23	-
	2	387	60.71	48.38	15.13	0.73	44.76	-1.6838	-1.6838	0.5746	8	21	19	20	1	17	18	2	23	22
	3	280	60.65	48.62	15.52	2.18	45.15	-2.4896	-2.4896	0.8054	8	21	19	20	1	17	18	16	22	15
	4	87	57.72	44.86	14.49	1.43	49.98	0.6503	1.1543	1.0547	19	8	21	23	2	1	22	17	18	-
	5	42	72.67	50.55	14.10	20.23	58.22	-3.1853	-3.1853	-2.2219	8	2	19	17	23	22	18	*	-	-
BI Total	1	517	60.18	143.41	23.62	0.10	4.70	0.8122	0.9848	0.4900	21	8	2	19	22	1	17	23	16	-
	2	387	60.71	143.23	23.45	0.05	5.84	-0.4553	-0.6707	-0.6544	21	8	17	2	19	1	20	23	22	*
	3	280	60.65	142.93	23.03	0.03	8.20	-0.3136	-0.4771	-0.4098	17	20	8	19	1	23	21	2	16	22
	4	87	57.72	143.41	22.99	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	42	72.67	151.05	21.19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VSMS	1	517	60.18	55.31	8.61	<0.02	10.50	0.0282	1.3311	1.0180	8	21	18	17	2	23	19	1	22	-
	2	387	60.71	55.50	8.33	<0.02	10.53	-0.0453	1.0584	0.8091	8	18	23	2	1	17	21	19	22	*
	3	280	60.65	55.58	8.68	0.02	12.00	0.2544	1.7072	1.7072	8	18	23	15	2	21	17	19	1	20
	4	87	57.72	53.99	13.84	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	42	72.67	59.17	10.41	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PSI 1	1	517	60.18	16.85	4.50	0.02	30.84	-0.3652	1.3733	1.3733	8	21	2	1	22	19	17	18	23	-
	2	387	60.71	16.87	4.47	0.24	32.17	-0.9735	0.7470	0.7143	8	21	20	2	19	17	1	23	18	*
	3	280	60.65	16.96	4.57	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	87	57.72	15.80	4.32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	42	72.67	16.21	4.45	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PSI 2	1	517	60.18	9.97	4.87	0.20	30.90	-1.0185	-1.0185	0.2480	8	21	1	22	2	19	18	23	17	-
	2	387	60.71	10.10	4.95	0.55	34.67	-1.4684	-1.4684	-0.0266	8	20	1	21	2	19	22	17	18	23
	3	280	60.65	10.33	5.16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	87	57.72	9.01	4.70	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	42	72.67	9.50	4.33	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PSI 3	1	517	60.18	8.43	3.70	0.38	38.92	-1.4151	1.4293	1.4293	8	19	21	1	23	17	22	*	-	-
	2	387	60.71	8.63	3.77	0.88	43.18	-1.8491	-1.8491	0.4382	8	19	21	1	17	22	23	20	18	2
	3	280	60.65	8.51	3.83	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	87	57.72	8.37	3.90	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	42	72.67	10.02	3.78	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PSI 4	1	517	60.18	12.58	3.87	0.26	36.39	-1.1721	1.7121	1.6779	8	19	21	2	1	18	17	23	22	-
	2	387	60.71	12.79	3.92	0.79	38.00	-1.7526	-1.7526	0.8771	8	19	21	2	17	20	1	18	22	23
	3	280	60.65	12.83	3.88	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	87	57.72	11.68	3.92	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	42	72.67	14.81	3.66	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PPVT IQ	1	517	60.18	84.77	18.81	0.10	10.61	0.7462	1.2325	1.2325	1	21	22	23	2	17	19	8	18	-
	2	387	60.71	85.10	19.15	<0.02	13.49	0.1795	0.8708	0.8708	1	20	21	17	23	19	2	22	18	8
	3	280	60.65	86.13	19.64	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	87	57.72	85.06	26.07	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	42	72.67	71.76	16.91	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
VSMS SQ	1	517	60.18	100.82	27.48	0.04	7.12	0.4941	0.8986	0.8986	21	2	8	17	19	1	23	22	18	-
	2	387	60.71	100.33	27.52	<0.02	9.09	0.1196	0.8680	0.8680	21	17	2	22	8	19	1	23	18	20
	3	280	60.65	100.11	27.77	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	87	57.72	91.49	47.90	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	42	72.67	93.78	22.33	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Notes: (1) An asterisk denotes analysis terminated before all variables were entered (see text).

(2) Variable names and numbers are as follows:

Number	Name
1	Race
2	Sex
6	Weeks
8	Age
15	PS-Exp.
16	P-Exp.
17	Hours
18	Center Size
19	Population
20	Father's Education
21	Mother's Education
22	Family Size
23	Income

(3) Variables Income, P-Exp., and PS-Exp. are omitted here for convenience. There were no significant betas for Income or P-Exp., and only one negative one for PS-Exp. for VSMS in Group 3.

F STEPWISE REGRESSION ANALYSIS BY TEST AND GROUP

otal % ariance or All pendent bles Used	T-Value for β Coefficient for Weeks			Order of Entry of Independent Variables[2]												Independent Variables with Significant β Coefficients in Final Multiple Regression Equation[3]									
	Starting	Maximum	Final	2	3	4	5	6	7	8	9	10	11	12	13	Race	Sex	Weeks	Age	Hours	C. Size	Pop.	M. Edu	F. Size	F. Ed.
1.73	-0.7075	0.7960	0.7960	8	1	21	22	17	23	2	19	18	-	-	-	-X	-	-	X	-	-	-	X	-	-
4.49	-1.0610	-1.0610	0.2878	8	1	20	21	17	19	23	2	18	22	-	-	-X	-	-	X	-	-	-	X	-	X
1.88	-0.8550	-1.3817	1.3574	8	20	1	21	15	16	17	2	19	18	22	23	-X	-	-	X	-	-	-	-	-	X
4.91	-1.6728	-1.6728	0.4734	19	8	22	23	21	1	17	18	2	-	-	-	-	-	-	X	-	-	-X	-	-X	-
5.15	-2.8390	-2.8390	-2.2613	8	2	19	17	23	22	18	*(1)	-	-	-	-	-	-	-X	-	X	-	-	-	-	-
1.68	-1.0982	1.3908	1.3572	8	21	1	19	2	18	22	17	23	-	-	-	-X	-	-	X	-	-	-X	X	-	-
4.76	-1.6838	-1.6838	0.5746	8	21	19	20	1	17	18	2	23	22	-	-	-X	-	-	X	-X	-	-X	X	-	X
5.15	-2.4896	-2.4896	0.8054	8	21	19	20	1	17	18	16	22	15	2	23	-X	-	-	X	-	-X	-X	X	-	X
9.98	0.6503	1.1543	1.0547	19	8	21	23	2	1	22	17	18	-	-	-	-	-	-	X	-	-	-X	X	-	-
8.22	-3.1853	-3.1853	-2.2219	8	2	19	17	23	22	18	*	-	-	-	-	-	-	-X	-	X	-	-	-	-	-
4.70	0.8122	0.9848	0.4900	21	8	2	19	22	1	17	23	18	-	-	-	-	X	-	X	-	-	-	X	-	-
5.84	-0.4553	-0.6707	-0.6544	21	8	17	2	19	1	20	23	22	*	-	-	-	-	-	X	-X	-	-	X	-	-
8.20	-0.3136	-0.4771	-0.4098	17	20	8	19	1	23	21	2	16	22	15	*	-	-	-	X	-	-	X	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.50	0.0282	1.3311	1.0180	8	21	18	17	2	23	19	1	22	-	-	-	-	-	-	X	-	-X	-	X	-	-
0.53	-0.0453	1.0584	0.8091	8	18	23	2	1	17	21	19	22	*	-	-	-	-	-	X	-	-X	-	-	-	-
2.00	0.2544	1.7072	1.7072	8	18	23	15	2	21	17	19	1	20	22	*	-	-	-	X	-	-X	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.84	-0.3652	1.3733	1.3733	8	21	2	1	22	19	17	18	23	-	-	-	-	X	-	X	-	-	-	X	-	-
2.17	-0.9735	0.7470	0.7143	8	21	20	2	19	17	1	23	18	*	-	-	-	X	-	X	-X	-	-	X	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.90	-1.0185	-1.0185	0.2480	8	21	1	22	2	19	18	23	17	-	-	-	-X	-	-	X	-	-	-	X	-X	-
4.67	-1.4684	-1.4684	-0.0266	8	20	1	21	2	19	22	17	18	23	-	-	-X	-	-	X	-	-	-	X	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.92	-1.4151	1.4293	1.4293	8	19	21	1	23	17	22	*	-	-	-	-	-X	-	-	X	-	-	-X	X	-	-
3.18	-1.8491	-1.8491	0.4382	8	19	21	1	17	22	23	20	18	2	-	-	-X	-	-	X	-X	-	-X	X	X	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.39	-1.1721	1.7121	1.6779	8	19	21	2	1	18	17	23	22	-	-	-	-X	X	-	X	-	-X	-X	X	-	-
.00	-1.7526	-1.7526	0.8771	8	19	21	2	17	20	1	18	22	23	-	-	-X	X	-	X	-X	-X	-X	X	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.61	0.7462	1.2325	1.2325	1	21	22	23	2	17	19	8	18	-	-	-	-X	-	-	-	-	-	-	X	-	-
.49	0.1795	0.8708	0.8708	1	20	21	17	23	19	2	22	18	8	-	-	-X	-	-	-	-	-	-	X	-	X
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7.12	0.4941	0.8986	0.8986	21	2	8	17	19	1	23	22	18	-	-	-	-	X	-	-X	-	-	-	X	-	-
.09	0.1196	0.8680	0.8680	21	17	2	22	8	19	1	23	18	20	-	-	-	-	-	-	-	-	-	X	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

ere no significant betas for Income or P-Exp., and only one negative one for PS-Exp. for largest t-value[1] when each variable was entered by itself into a regression equation with the variables from the preceding step. For example, in the analysis of the PPVT RS (raw score) with Group 1, Variable 1 (Age) was the first variable to enter after Variable 0 because, of the nine variables considered, it had the largest beta coefficient t-value (and therefore F-ratio, since $t^2 = F$) in a 2-variable regression equation employing weeks. Variable 3 (Race) was selected next because, of the remaining eight variables, it had the largest beta weight t-value in a 3-variable regression equation employing weeks and age. An asterisk after the last variable entered means that some remaining variable or variables never entered the analysis at all since the criterion values for selection were too low.

Finally, for each analysis, an indication is provided of those variables whose final beta coefficient had a t-value corresponding to a probability of 0.05 or less. In addition, if the beta coefficient was negative, this is indicated with a minus sign.

For ease of reference, the mean and standard deviation of each independent variable for each group is presented in Exhibit 46.

c. Intertest Relationships

One way of addressing the question of whether the tests that were used provide different information about Head Start children is to examine the intertest correlations. Tests that are highly intercorrelated do not add unique information about the characteristics or performance of the children.

From the multivariate By-Center covariance analysis, we obtained the intertest product-moment correlation coefficients shown on the lower left half of the matrix in Exhibit 47. Product-moment coefficients from the multiple regression analysis are shown in the upper right half of the exhibit. Thus, for example, based on center average scores, the PPVT raw scores correlated 0.86 with the PSI total raw scores in the covariance analysis. These intercorrelations are of

---

[1] Actually, the selection criterion was an F-value; however, $F = t^2$.

EXHIBIT 46 - MEANS AND STANDARD DEVIATIONS OF INDEPENDENT VARIABLES BY GROUP

Variable		Group One	Group Two	Group Three	Group Four	Group Five
1. Age in Months	$\bar{X}$	60.18	60.71	60.65	57.72	72.67
	σ	8.71	8.67	8.27	7.01	7.94
2. Sex (Proportion of Males)	$\bar{X}$	0.49	0.49	0.50	0.46	0.45
	σ	0.50	0.50	0.50	0.50	0.50
3. Race (Proportion of Non-Whites)	$\bar{X}$	0.58	0.53	0.48	0.62	0.00
	σ	0.52	0.53	0.54	0.49	0.00
4. Weeks	$\bar{X}$	19.90	19.86	21.20	17.66	12.79
	σ	7.45	7.15	6.99	5.21	7.12
5. Hours Per Day	$\bar{X}$	3.79	3.79	3.90	4.25	3.31
	σ	1.32	1.28	1.41	1.48	1.14
6. Center Size (Number of Classes)	$\bar{X}$	2.61	2.60	2.80	2.83	4.81
	σ	1.43	1.51	1.56	1.50	2.89
7. Population (Log of Town Size)	$\bar{X}$	5.14	5.01	5.16	5.44	5.14
	σ	1.22	1.27	1.18	1.31	0.76
8. Mother's Education	$\bar{X}$	5.21	5.25	5.28	5.02	3.36
	σ	1.11	1.10	1.06	0.99	1.56
9. Father's Education	$\bar{X}$	-	5.02	5.05	-	-
	σ	-	1.31	1.30	-	-
10. Family Size	$\bar{X}$	6.69	7.11	6.96	6.23	7.02
	σ	2.25	2.51	2.47	2.42	1.69
11. Income	$\bar{X}$	3,820	4,270	4,220	3,600	3,180
	σ	2,110	2,060	1,950	2,120	2,010
12. PS-Exp.	$\bar{X}$	-	-	2.54	-	-
	σ	-	-	0.92	-	-
13. P-Exp.	$\bar{X}$	-	-	2.27	-	-
	σ	-	-	0.84	-	-

course contaminated or inflated by the intercorrelation of other variables associated with center average scores.

EXHIBIT 47 - INTERTEST PRODUCT-MOMENT CORRELATION COEFFICIENTS FROM BY-CENTER MULTIVARIATE COVARIANCE ANALYSIS AND MULTIPLE REGRESSION ANALYSIS

	PPVT	PSI	BI	VSMS
PPVT		0.89	0.14	0.29
PSI	0.86		0.15	0.33
BI	0.23	0.18		0.15
VSMS	0.36	0.41	0.08	

From By-Center Regression Analysis
From By-Center Covariance Analysis

In Exhibit 48 are shown the least squares estimates of the intertest covariances based on center averages (with the effects of age, sex, race, etc., removed) which were used along with the residual variances to estimate the residual correlation coefficients; these are given in the upper right half of the matrix on the basis of the regression model and in the lower left half on the basis of the analysis of covariance model.

EXHIBIT 48 - INTERTEST RESIDUAL CORRELATIONS

	PPVT	PSI	BI	VSMS
PPVT		0.75	0.25	0.17
PSI	0.70		0.26	0.18
BI	0.18	0.12		0.12
VSMS	0.14	0.18	0.03	

From By-Center Regression Analysis
From By-Center Covariance Analysis

Thus, only the PPVT and PSI have much common variance. In effect, they share around 50 to 55 percent of their variance. The other two instruments, which are based on information about children provided by teachers or aides, have relatively little variance in common with each other or with the individual tests.

The loss in intertest correlation that occurs when the effects of various independent variables are partialled out is readily understandable when the partial correlation coefficients of the independent variables with each of the tests are examined. From the By-Center multivariate regression analysis, they are shown in Exhibit 49.

EXHIBIT 49 - PARTIAL CORRELATION COEFFICIENTS OF THE INDEPENDENT VARIABLES WITH EACH OF THE FOUR TESTS

	Independent Variables					
Tests	Age	Sex	Race	Population	Poverty	Exposure
PPVT	0.59	0.07	-0.35	-0.06	-0.20	-0.09
PSI	0.74	-0.08	-0.21	-0.15	-0.13	-0.10
BI	0.17	0.02	-0.05	0.26	0.06	-0.08
VSMS	0.34	-0.08	0.12	0.01	-0.13	0.06

The intercorrelation of the BI with the other tests was relatively unaffected by removing the effects of the independent variables, since it had little correlation with these variables. Removing the age variable from the intercorrelation of the VSMS with the PPVT and the PSI, on the other hand, reduced their intercorrelation markedly, since the VSMS has a sizeable correlation with age.[1] On the other hand, the PPVT and PSI show a high intercorrelation even with the age variable removed, suggesting that they do indeed measure similar functions or processes, or at least that one is a fairly good predicator of the other.

---

[1] The correlations of test scores with age are probably themselves somewhat inflated because of the range of ages included for analysis.

The Pre-School Inventory subtests were not examined in the By-Center analyses; that was done only in the By-Child analyses. To give an indication of the observed relationships of the PSI subtests to each other and to other instruments, residual correlation coefficients were calculated for Group 1 (N = 517). The correlations eliminate the intercorrelations with the 10 independent variables. The residual correlation coefficient estimates are shown in the upper right half of the matrix in Exhibit 50; for comparison, the product-moment correlation coefficients for the same data are given in the bottom left half of the exhibit.

EXHIBIT 50 - RESIDUAL CORRELATION COEFFICIENTS ESTIMATES AND PRODUCT-MOMENT CORRELATION COEFFICIENTS FOR PSI SUBTESTS AND OTHER TESTS FOR GROUP 1 (N = 517)

	PPVT	PSI_T	BI	VSMS	PSI_1	PSI_2	PSI_3	PSI_4
PPVT		0.67	0.34	0.22	0.54	0.61	0.49	0.54
PSI_T	0.78		0.45	0.30	0.84	0.84	0.78	0.77
BI	0.34	0.42		0.31	0.37	0.37	0.35	0.40
VSMS	0.32	0.40	0.34		0.40	0.25	0.29	0.30
PSI_1	0.67	0.89	0.39	0.34		0.64	0.54	0.55
PSI_2	0.73	0.89	0.37	0.34	0.74		0.56	0.52
PSI_3	0.66	0.87	0.32	0.33	0.68	0.70		0.51
PSI_4	0.68	0.86	0.38	0.40	0.70	0.65	0.73	

Estimates of Partial Correlation Coefficients
Product-Moment Correlation Coefficients

These coefficients in Exhibit 50 are based on a different model and sample from the By-Center coefficients. They include in the coefficients whatever variations there are in intertest correlation from center to center. For convenience, the product-moment and partial correlation coefficients of the 10 independent variables with each dependent variable for Group 1 are given in Exhibit 51.

In terms of magnitude of relationships between the coefficients within and between models (By-Child and By-Center), the following statements are appropriate:

- By-Child coefficients of both kinds (product-moment and partial) are typically larger than the corresponding By-Center coefficients of either kind.
- Product-moment correlations within models are typically (in 80 percent of the cases) larger than the corresponding partial correlation coefficients. (Exceptions in all but one case involve the BI.)

Thus, the partial correlation coefficients of the By-Center model appear generally to give the most conservative or minimal estimate of intertest correlations.

EXHIBIT 51 - CORRELATION OF INDEPENDENT AND DEPENDENT VARIABLES FOR GROUP 1

Test		Age	Sex	Race	Weeks	Hours	Center Size	Population	Mother's Education	Family Size	Income
PPVT	PM(1)	0.48	-0.01	-0.34	-0.03	-0.09	-0.09	-0.30	0.04	0.01	0.05
	PC(2)	0.41	-0.05	-0.24	0.04	-0.06	-0.02	-0.04	0.19	-0.07	0.05
PSI	PM	0.59	0.08	-0.30	-0.05	-0.09	-0.14	-0.38	0.03	0.04	0.04
	PC	0.51	0.06	-0.14	0.06	-0.04	-0.07	-0.13	0.22	-0.06	0.04
BI	PM	0.09	0.11	-0.02	0.04	-0.03	-0.02	0.04	0.11	-0.06	0.04
	PC	0.12	0.10	-0.04	0.02	-0.03	-0.02	0.07	0.12	-0.06	0.02
VSMS	PM	0.25	0.07	-0.04	0.00	-0.06	-0.16	-0.06	0.07	0.02	0.06
	PC	0.23	0.05	0.01	0.05	-0.05	-0.14	0.03	0.12	-0.01	0.04
PSI_1	PM	0.51	0.15	-0.20	-0.02	-0.08	-0.10	-0.28	0.03	-0.01	-0.01
	PC	0.45	0.14	-0.06	0.06	-0.04	-0.03	-0.06	0.17	-0.07	-0.01
PSI_2	PM	0.50	-0.03	-0.26	-0.04	-0.03	-0.11	-0.30	0.05	-0.03	0.02
	PC	0.44	-0.08	-0.14	0.01	0.02	-0.03	-0.07	0.19	-0.12	0.02
PSI_3	PM	0.56	0.03	-0.33	-0.06	-0.09	-0.14	-0.42	0.00	0.11	0.07
	PC	0.44	0.00	-0.15	0.06	-0.05	-0.07	-0.18	0.17	0.03	0.07
PSI_4	PM	0.53	0.15	-0.29	-0.05	-0.12	-0.16	-0.37	0.01	0.09	0.06
	PC	0.43	0.15	-0.12	0.07	-0.08	-0.10	-0.15	0.18	0.02	0.06

Notes: (1) PM denotes product-moment correlation.

(2) PC denotes partial correlation.

V. DISCUSSION

This section will discuss the results reported in the preceding section. There are three major areas of consideration: (1) the main hypothesis, (2) alternative hypotheses, and (3) comments on exceptions to the main results.

A. The Main Hypothesis

The principal result of this study may be stated as follows: there was no statistically reliable evidence of a treatment effect observed for the main samples of children tested, based on measures utilizing total test raw scores. Neither a systematic gain with time nor loss with time was found, nor was there evidence of a significant but non-linear relationship.

This conclusion rests on four criterion statistics:

- Wilks' Lambda for the covariance analyses.
- The F-ratios for individual instruments in the covariance analyses.
- The t-values (or confidence intervals) for the β coefficient for the exposure variable in the regression analyses.
- Proportions of test variance attributable to exposure time.

The covariance analyses provide the most powerful bases for statements about treatment effects. It is true that the By-Child analyses of covariance are of limited value, since the universe of these subgroups would be of indeterminate generality and of limited operational interest even if there had not been losses (sometimes severe) owing to missing family data. Furthermore, the discriminating power of many of the By-Child covariance analyses is rather low, owing to the small sample sizes. However, the primary By-Center covariance analyses, whether for the total eligible sample of English-tested children or for the total eligible sample, can be interpreted with more confidence. Nevertheless, for main and subsamples, the result is negative. This lack of evidence of a relationship between program duration and performance (or rating) of child applies whether the measures are considered jointly or (except in one case for the Behavior Inventory) individually.

The regression analyses do not bear directly on the question of a treatment effect, since the treatment or exposure variable is not kept independent of other variables in these analyses. However, to the extent that the β coefficient for the exposure variable is significantly different from zero, or has a confidence interval that does not include zero, there is presumption of an effect. Only one group--the eligible children tested in Spanish--had an exposure variable (Weeks) with a significant β coefficient, and it was negative for the two measures analyzed (the PPVT and PSI). Furthermore, except for the group of Spanish-speaking children, the percentage of variance in the test scores accounted for by the exposure variable was almost never more than 2 percent, and, in the By-Child regression analyses it was typically less than 1 percent on the basis of our procedure.

As described in subsection IV. B, the attempt was made to replace Weeks or program length with the total number of presumed hours in the program as a measure of exposure. In that case, too, the β *weight for exposure was not significant.*

These considerations lead to two major questions that warrant further comment:

1. Assuming that Head Start programs can produce measurable cognitive, social and emotional effects in children, what factors or hypotheses may account for the lack of evidence for them in the present study?

2. How can those few cases of significance that did occur be accounted for?

These two questions are discussed in that order below.

B. Alternative Hypotheses

What may have obscured our observation of treatment effects? The question can be addressed by consideration of a number of alternative hypotheses.

It should be made clear from the outset that any discussion of effects involves only consequences measurable by the tests employed. We are not concerned with medical or dental effects, for example, although these may well have been manifold. Let us also clarify what is meant by an effect, or at least the evidence for one. We cannot speak of gain scores, or changes in a child's score from the beginning to the end of the program. As the experiment was designed, the Short-term (S) group is the control or comparison group. The primary evidence for an effect is a difference between the means of S and either or both Medium-term (M) or Long-term (L) samples for the dependent variables considered together or individually. There are certainly other indicators, as noted above, although they may be less compelling, both in terms of power and operational significance. For example, there may be shifts in variability with no change in means. There may be consistent relationships between the means or medians of the samples (e.g., $L > M > S$) even though the differences are not significant by a parametric test. There may be β weights in a multiple regression equation that are significantly different from zero, lending credence to the hypothesis of an effect associated with a treatment variable. However measured, whenever a treatment level distinction (S, M, L) is retained, an effect in this study is basically a score change relative to the S level. We have avoided identifying effects with a particular psychological construct or product. For the time being, an effect is defined purely operationally.

With this background, let us consider possible obscuring factors.

1. Non-Comparability of Samples on Essential Uncontrolled Variables

The most compelling of these variables would be starting level of performance. There is no direct rejoinder to the hypothesis that L children, for example, had lower starting levels at the beginning of the program than the S children. (There is no evidence to support that hypothesis, either.) Our covariance adjustments for age, sex, race, etc., in effect made the three samples uniform with respect to those variables. However, demonstrable equalization of entrance abilities or achievements could only have been accomplished by pre-testing or by randomization of assignment of children to treatments. Of these two methods, the latter would have been the more satisfactory and the less possible. (Actually, of course, neither method was possible.)

No argument concerning the actual or statistical matching of control variables such as age, sex, number of siblings, mother's educational level, family income, and so on, can directly resolve the question of entry comparability of the children in this study. However, the logic of the argument for non-comparability is slippery. If, for example, there were a systematic incremental effect of the program, then our results suggest that there would have to have been a progressive decrease in entry level as a function of length of program. It is hard to imagine what would produce such a selection process.

Selective attrition during programs could have produced the same result. An attempt was made to ascertain whether there was some correlation between the length of the program, defined by the average length of attendance in weeks of a center's sample of children, and the retention rate of the center. Retention rate is defined as the ratio of the number of children attending the program at the time of the testing who had been in the program at its start to the total number of children registered at the commencement of the center's program.

There were great difficulties in obtaining the data necessary for these measures in many cases, owing to various administrative practices of the programs. Although there appeared to be some positive relationship in the rank order correlation between these two variables, it was not significant ($r_s = 0.16$; $t = 1.33$; $df = 67$; $p > 0.05$). Thus, available evidence does not generally support the hypothesis of a relationship between program duration and attrition.

Another selection bias which could have had the same effect as experimental mortality was the absence of children, otherwise in attendance, during the week of testing. There was one case, for example, in which the center was more than decimated by an outburst of mumps and chickenpox while the tester was there. However, we have no practical way of addressing this selection hypothesis.

In sum, we can neither accept nor reject this possible obscuring factor. It was hoped that the sampling plans used would minimize such a sampling bias; we cannot, however, demonstrate by data or design its non-existence.

2. Inadequate Sample Sizes

It is entirely possible that there were incremental differences between S, M, and L samples that did not show up owing to the limited number of centers and children. There is little doubt that sufficiently large samples will generally produce significant differences. However, the differences are apt to be so small as to be operationally meaningless.

To provide an indication of the magnitude of differences we could have expected to find, the simultaneous 95 percent confidence limits for differences between adjusted means for the PPVT and the PSI were calculated. These intervals, shown in Exhibit 52, are listed for the main By-Center group, and for the By-Child groups with the smallest (3-year-old non-white female) and largest (4-year-old non-white male) total sample sizes.

From Exhibit 52, it can be seen that, for example, we could reasonably expect to detect L-S PPVT raw score differences of more than about 4 points (the half-interval length), or L-S PSI differences of about 4-2/3 points or more. The sensitivity of the observations decreases drastically, of course, with a reduction in sample size, to the point (exemplified by the smallest By-Child group) at which differences that could have been detected with assurance are so big as to be meaningless.

The intervals shown in Exhibit 52 were, as stated above, calculated for all contrasts taken together. Thus, they are conservative. Confidence intervals for the difference between one pair of means taken individually (e.g., the L-S PPVT difference) can be obtained by dividing

EXHIBIT 52 - SIMULTANEOUS NINETY-FIVE PERCENT CONFIDENCE INTERVALS FOR DIFFERENCES BETWEEN ADJUSTED TREATMENT MEANS

Group	Type of Analysis	Test	Treatment Level Difference		
			L-S	M-S	L-M
Eligible Tested in English	By-Center	PPVT	1.31 ± 3.91	0.54 ± 3.64	0.77 ± 4.67
		PSI	0.48 ± 4.67	0.32 ± 3.89	0.16 ± 5.55
4-Year-Old Non-White Males	By-Child	PPVT	-0.17 ± 6.76	-1.18 ± 9.58	1.01 ± 10.00
		PSI	-0.28 ± 7.88	-0.66 ± 11.17	0.38 ± 11.67
3-Year-Old Non-White Females	By-Child	PPVT	47.60 ± 74.50	28.65 ± 74.02	18.95 ± 35.82
		PSI	7.64 ± 34.57	3.53 ± 34.36	4.11 ± 16.63

the $k\sigma$ parameter by $\sqrt{2}\left(\frac{\sqrt{F^{.05}_{2,v}}}{t^{.05}_{v}}\right)$. Thus, for the L-S PPVT comparison we should expect to detect differences larger than approximately 3.1 points. Similarly, with the L-S PSI comparison, we should anticipate differences at the 0.05 level that are greater than about 3.7 points. The real issue is not basically whether the sample size was adequate; the issue is the meaning of adequate--i.e., what magnitude of difference it is meaningful to try to detect. It is not possible to take a stand on this issue here. However, the ability to stipulate goals in this form surely should be one of the major objectives of educational and psychological research efforts.

3. Non-Uniform Treatments and Effects Among or Within Samples

Obviously, teachers, children, procedures, and facilities or environment varied from class to class as well as from CDC to CDC. Objectives and goals probably showed similar diversity. Consequently, whatever the specific nature of a Head Start center program, the "treatment" perforce varied at least from classroom to classroom. The results of the treatment undoubtedly varied from child to child. The issue is why the net result of the infinite specific and different treatments would be zero.

The question is complex, and clearly unanswerable without recourse to specifics about treatments and effects--a specificity far beyond the available data in this study. Indeed, the specification and measurement of treatments in child development and in education is still one of the most difficult problems in educational research.[1] Particularly in pre-school programs, the treatment variables are difficult to define except on a gross level and more difficult to measure reliably.

---

[1] See Gage, N. L. (ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963, passim.

There are undoubtedly interactions of treatments and subjects, however treatment is defined. Whether we think of variables as teachers or as center programs (or both), different levels (individually or in combination) very likely interact with pupil variables. Since we did not (indeed, could not) design our sample to provide observations covering a known range of teacher-program-child variables, it could be argued that there were no apparent effects because there was an inadequate number of "positive" or "optimum" pupil-treatment observations.[1] Perhaps if we had used classrooms as the basic sampling unit, and had tested all children in each sample classroom, we would have had sufficient numbers of specific teacher-program-pupil combinations at different levels to have seen an effect. The net effect would presumably result from the occurrence of the differential effects. The theory and observation of a teacher (and/or program) - child interaction is considered fundamental not only in education,[2] but in related enterprises such as psychotherapy.[3]

The problem here is not whether there are treatment-subject interactions, but what the net effect of treatments may be. For the design of this study, an effect, measured for instance by a difference, Δ, in S and L means (i.e., $\Delta = L - S|$), can have one of three values:

Δ is positive.

Δ is negative.

Δ is, in effect, zero.

---

[1] Validity and sensitivity of the instruments are assumed for purposes of this argument.

[2] See Gage (*op. cit.*) passim.

[3] See, for instance, Kiesler, Donald J., "Some Myths of Psychotherapy Research and the Search for a Paradigm," *Psychological Bulletin*, 1966, Volume 65, pp. 110-136.

Let us assume, for purposes of discussion, that the experimental groups were comparable in starting level and that the two tests (PPVT and PSI) are each appropriate for measuring cognitive status or achievement level of the children. The measures themselves depend upon the responses of the subject to some partially known stimulus. The responses are binary (from the experimenter's point of view, a response is either right or wrong), and there is the usual confounding of cognition and motivation, which (in highly oversimplified and non-operational terms) is something like this:

	Motivation	
Cognition	Willing to Respond	Unwilling to Respond
Able to Respond	Correct Response	No Response or Wrong Response
Unable to Respond	Wrong Response or No Response	No Response

Here, the term "able" means "has the requisite knowledge and skills," while "willing" means "attempts to make best response possible according to perceived requirements (or rules) of the game (test)." As the relationships are depicted here, only the correct response has unambiguous meaning if we ignore the role of chance. In this study there were no specific criteria of motivation beyond the finding of the tester that the child was "testable" in the sense that he would stay in the situation and respond at all.

The effect of participation in a Head Start program could be in either or both realms and produce the same result. Conceptually, however, the action of a treatment may differ according to the area or realm of effect. That is, it is reasonable to conceive of cognitive effects as having a zero point and increasing in magnitude, complexity, scope, etc. It is more difficult to conceive of the cognitive effects as bi-polar, with changes (losses) occurring as a result of participation in a program. On the other hand, motivational effects could easily be positive or negative, and could interact with cognitive effects in a variety of ways, uniformly or selectively (i.e., in terms of individuals or subgroups). However, regardless of what sort of treatment-pupil interaction is assumed, and regardless of what sorts of intereffect combinations are assumed, the absence of an observed treatment effect (i.e., a difference in means) suggests that:

- There was no measurable effect in the cognitive or cognitive-motivational realms.[1]
- Positive effects in the cognitive area were nullified by opposing effects in the motivational area.
- Positive effects in the motivational area were nullified by negative effects in the cognitive area.

Another alternative, that positive effects in the motivational area were not accompanied by positive effects in the cognitive area, is tenable if, for example, it is assumed that S, M, and L samples were systematically less developed in the cognitive and that the positive motivational effect simply maximized the use of otherwise unaffected knowledge and skills.

Any of the above hypotheses is possible, regardless of whether assumptions of uniformity or diversity of treatments and subjects are made. If one accepts the interaction point of view--that the major effects depend on the interaction of subjects and treatments--PRC's results suggest either (1) that there were too few optimum combinations to make a measurable difference (a rareness or scarcity that does not bode well from the point of view of matching teacher selection, training, program structure and content, teacher behavior, or whatever manipulable variables are considered the effective dimensions of a treatment with the appropriate pupil variables), or (2) that there were as many and as strong negative combinations of treatments and subjects as there were positive ones.

---

[1] This hypothesis in no way rules out the possibility of major effects occurring with all or many children early in their participation in a program (e.g., during the first 1 to 6 weeks).

4. Inappropriate Instruments or Tests

One major reason that effects which actually existed may not have been noted centers around the measuring instruments. The instruments may not have been appropriate for a number of technical and conceptual reasons.

The study was not designed to test hypotheses about the extent to which the tests were suitable or appropriate. However, the question of instrumentation is a major one in Head Start evaluation, as in other educational and social programs; it is thus appropriate to make what comments we can in this area. The tests will be discussed in terms of validity and reliability, not so much in the usual manner of providing coefficients of systematically collected correlational data, but by drawing on the data and inferences reasonably bearing on these properties.

a. Validity

From a program point of view, validity depends in part on objectives and goals. Head Start, like many compensatory programs, is careful not to impose national goals which are too tightly defined. The following remarks, taken from a report submitted to the House Education and Labor Committee in 1966,[1] illustrates the variety implicit in Head Start programs:

> THE NATURE OF A HEAD START PROGRAM
>
> Head Start Programs come in all sizes and shapes. Some are three hours a day--others for a full day. The great majority of Head Start children have, thus far, been enrolled in SUMMER programs. This report is focused on the ways in which Head Start might be expanded to FULL-YEAR programs. Some of these would operate nine months a year--others up to twelve.

---

[1] Report submitted by Bernard L. Boutin, Acting Director, Community Action Program, Office of Economic Opportunity, 29 March 1966. Printed in: Supplemental Appropriation Bill, 1967: Hearings before Subcommittees of the Committee on Appropriations, House of Representatives, 89th Congress, pp. 141-156.

> Local needs and local decisions are the determining factors in developing and operating the program with one important exception. The Economic Opportunity Act requires that programs show promise of making a meaningful contribution to the elimination of poverty. To this end OEO insists that all programs must contain those elements of quality which are believed essential to really helping the child and his family. Programs devoid of content and purpose, programs aimed solely at permitting parents to work, and programs concerned exclusively with educational development cannot be funded under Head Start for none of these types shows sufficient promise of helping to end poverty.
>
> Head Start programs are carried out in Child Development Centers which:
>
>> Provide help to both the child and his family. There must be the widest possible opportunities for parents to participate in program decisions and operations and to themselves be beneficiaries of the programs.
>>
>> Offer a comprehensive range of services which are critical to the child's development including medical and dental care, social services, nutritional support and a well designed program of daily activities designed to meet the needs of the individual child.
>>
>> Are a true community project involving cooperation among the professional and non-professional staffs, parents of the children, professional individuals and organizations serving children, government agencies, and ordinary citizens--young and old--who are able and willing to volunteer their time and skills.
>
> The particular methods and activities needed to accomplish these objectives cannot be specified by OEO. It can, and does, do everything in its power to advise communities of methods and activities which have shown promise in other communities.

It is clear from this statement that the educational and developmental goals of individual programs are not fixed. There is no indicated requirement to improve communication skills, develop school readiness, improve social skills or responsiveness, or foster any specific behavioral change. Nevertheless, the programs have typically an educational-developmental tone or appearance. Consequently, measures applied to program evaluation ought to include variables related to these areas of endeavor and/or behavior. The PSI was designed to have content validity with respect to achievement expected by teachers of children entering school. The VSMS ideally should have content validity for some aspects of personal-social development. The BI contains a number of descriptive scales that appear to suggest the kind and direction of behavior which teachers probably like to encourage. The PPVT presumably taps a set of skills involved in verbal communications. From a program point of view, then, the tests at least seem to have relevance, or face validity.

One criterion of validity is sensitivity of scores to those variables to which one would normally expect them to be sensitive. For example, the PPVT, PSI, and VSMS should all be sensitive to age, and they are. The correlation of these tests with age was higher than with any other variable (see Exhibits 49 and 51), and the beta weight for age was significant regardless of subgroup (see Exhibit 45). The individual test scores (PPVT and PSI) correlate with race, with whites tending to score higher than Negroes, although the coefficients are not as high as for age. All four measures appear to be sensitive to the level of education of the child's mother, as would be expected, and the PPVT and PSI are related also to the level of the father's education (see Exhibit 45). However, the BI and VSMS appear to be entirely unrelated to that latter variable. There is some indication that the tests (other than the VSMS) are slightly correlated negatively with family size, as might be expected.[1] While the beta weight for poverty is not significant in the By-Center regression analysis (see Exhibit 41), the signs of the partial correlations of the PPVT, PSI, and VSMS are at least in the right or expected direction (see Exhibit 49), as are the signs of the coefficients for income and family size in the By-Child regression analysis (see Exhibit 51). There is finally some slight indication of sensitivity of test scores to sex, although it appears somewhat complex (see, for example, Exhibit 51). Thus, at least the PPVT, PSI, and VSMS appear to be somewhat sensitive to variables to which one would expect them to be sensitive.[1]

---

[1] Anastasi, Anne and Rita Y. D'Angelo, "A Comparison of Negro and White Preschool Children in Language Development and Goodenough Draw-a-Man IQ, Journal of Genetic Psychology. Vol. 81, 1952, pp. 147-165.

The BI generally appears to correlate little with any of the variables studied. There is indication of some positive correlation with age. There is yet slimmer evidence of a correlation with sex, although boys tended to be rated slightly higher than girls. In the case of the By-Center multivariate regression analysis, the beta coefficient for population in the BI equation was significant and positive (Exhibits 40 and 41). However, the variables to which the BI should be sensitive are not immediately apparent.

Both the VSMS and the BI depended on information provided by the teacher or teacher aide about the child. Thus, there is an additional problem of validity in these instruments.[2] We have no independent criterion or measure of validity of teachers' ratings of children on BI scales. We do know that many teachers found the BI difficult to complete, and that some did so hastily, while others deliberated each scale at length. In VSMS interviews, the testers sometimes found teachers indicating that the children were much alike in abilities, or stating ignorance of much of the child's capabilities. In a few cases, testers noticed that an individual child could not in fact do what the teacher had said that he could. However, at least in Vineland interviews testers were able to make some checks on the probable authenticity of the information.

One indication, however, that the BI may be somewhat sensitive to actual behavior of the children (as seen, interpreted, converted, and rated by teachers), comes from an attempt to examine stereotypes. We asked 15 teachers of the deaf in a summer course to imagine and rate on BI's:

---

[1] In interpreting the various correlations, it is necessary to bear in mind that Head Start children are a selected sample whose range of variation on a number of variables is restricted.

[2] See Appendix D, Tests.

- A disadvantaged 5-year-old boy.
- A non-disadvantaged 5-year-old boy.
- A disadvantaged 5-year-old girl.
- A non-disadvantaged 5-year-old girl.

The mean "scores" are shown below in Exhibit 53.

EXHIBIT 53 - MEAN BEHAVIOR INVENTORY SCORES FOR HYPOTHETICAL CHILDREN

	Non-Disadvantaged	Disadvantaged
Boy	152.80	109.47
Girl	149.53	114.80

Where the mean scores for the imagined disadvantaged children tended to favor the girls, the obtained BI scores tended to be higher for boys. More striking, however, is the comparison of the means of the hypothetical children with the obtained means for 5-year-old males and females (see Exhibits 21A, 21B, and 21C). The means for the imagined non-disadvantaged children are exactly in the range of obtained unadjusted means for 5-year-olds, as the exhibits show. In no case did any of the S, M, or L 5-year-old groups have a BI mean as low as 115. This suggests that, if the hypothetical scores reflect stereotypes of some generality, Head Start teachers may have been more discriminating and rated the children more consistently with their actual behavior.

A similar argument comes from a comparison of BI scores obtained in the summer 1965 Head Start research evaluation effort.[1] For the sample obtained, the post-test mean for 5-year-olds was 127.82.

---

[1] Planning Research Corporation, PRC R-795, Results of the Summer 1965 Project Head Start, Volume I, Cort, H. Russell, Jr., et al, 9 May 1966. (See Exhibits I-14 and I-15.)

While that is somewhat lower than most of the obtained BI means for 5-year-olds in this study, it is still higher than the hypothetical child means.

No data bearing on the reliability of the BI is available. We did not do a split-half or odd-even correlation analysis of it, nor did we have the opportunity to do a re-test. A factor analysis of the BI was performed, using the nine subtests as variables; however, the results were not informative and will not be discussed here.

The raw scores from all instruments were examined for tester effects. The hypothesis that the tester affects the score on a test was tested by a one-way analysis of variance of individual test raw scores for each of four age levels. The treatment variable was the tester. Results for all instruments, including the DAP, are shown in Exhibit 54. It is apparent that the BI was the only test with no evidence of a tester effect, as would be expected. This is not, of course, positive evidence of the validity or appropriate sensitivity of the BI, but at least there was no strange negative evidence.

The implications of Exhibit 54 for the other tests, and indeed for the findings of the study, bear some comment. First, does the result suggest a reason for failure to detect an effect? The exhibit lists the code number of the testers associated with highest and with the lowest means in each cell where the F-ratio was significant. This gives some indication of the extent to which testers were consistent in producing an effect, however it may be regarded. As far as the PPVT, PSI, and DAP are concerned, the conclusion is that one tester (05) appears to be associated with low scoring groups. However, there are two points to be noted. These are that the low PPVT and PSI samples of 3-year-olds were the same children and that the sample size was, in fact, two. Furthermore, a number of the 23 low DAP 5-year-olds associated with this tester were Spanish-speaking children. These appear to be sufficient reasons for disregarding the possibility that this tester was biasing the results substantially.

EXHIBIT 54 - RESULTS OF ANALYSIS OF VARIANCE OF TEST SCORES, BY TESTER AND AGE OF CHILDREN TESTED(1)

Age	PPVT	PSI	BI	VSMS	DAP	Total Number of Testers
3	Tester 19(2) Tester 05(2)	Tester 15(4) Tester 05(4)	NS	NS	NS	11
4	NS(3)	Tester 13(4) Tester 19(4)	NS	Tester 04(4) Tester 15(4)	Tester 17(4) Tester 11(4)	16
5	NS	Tester 16(4) Tester 18(4)	NS	Tester 04(4) Tester 15(4)	Tester 12(4) Tester 05(4)	18
6	NS	NS	NS	Tester 03(2) Tester 08(2)	NS	6

Notes: (1) For age-test cells where significant score effects were found between testers, testers' identification numbers have been indicated. A tester's number in the top part of a cell indicates the tester whose subjects obtained the highest scores for a given age group, and a tester's number in the lower part of a cell indicates the tester whose subjects obtained the lowest scores for a given age group.

(2) Significant at 0.05 level of confidence (loc)

(3) NS = not significant

(4) Significant at 0.01 level of confidence (loc)

With respect to the VSMS, there appear to be two testers associated with high and low scores. Sample sizes range from 17 to 49. The tester "effect," in this case, could not be on the behavior of the child but on the responses of the teacher. The specific distribution of children in these high or low groups has not been examined with respect to S, M, or L. However, Tester 04 tested in M and L centers, while Tester 15 tested in all three types. It seems reasonable that if Tester 04 was biasing results, it was in favor of a "treatment" effect, whereas Tester 15 would tend simply to subtract a constant from each level. This is speculation, of course.

With respect to the PPVT, the high group tester tested five 3-year-olds in an L center. The six 4-year-olds who formed this tester's low group on the PSI were in the same center. The other high testers on the PSI tested in M centers (Tester 13), S and M centers (Tester 16), and M and L centers (Tester 18). Sample sizes of children range from 2 (Tester 05) to 28. With the exception of Tester 19, in no case were the extreme groups the total number of children tested by a given tester.

The overall impression is that while there is evidence for the interaction of testers and children,[1] it does not seem systematically related to particular tester for the PPVT or PSI, nor to S, M, and/or L for any of the measures.

Another criterion of validity of the tests is the extent to which they are related to other indicators or measures of performance. A very limited opportunity to investigate this possibility was provided by data from the Paid and Volunteer Worker's Evaluation Form collected from the staff of the centers by the Bureau of the Census. The form consists of a series of rating scales to be checked by the worker. One set of scales relates to changes in the children. The respondent is asked to rate as "much better," "better," "no change," "worse," or "much worse" the following:

---

[1] Evidence that the sex of the DAP drawing was related to the sex of the tester for this sample is reported by Datta, Lois-Ellin and Drake, Ann, "Objective Sexual Differentiation in the Drawings of Preschool Children," Paper submitted to Journal of Consulting Psychology. (Mimeographed copies are available from Dr. Datta, 2N256, Building 10, National Institutes of Health, Bethesda, Md. 20014).

I feel that in general children attending the Operation Head Start program were changed in the following ways:

1. Getting along with other children.
2. Self-confidence.
3. Speaking ability.
4. Everyday manners.
5. Finishing what he starts.
6. Doing what he's told.
7. Interested in new things.
8. Can do things on his own.
9. Chances of success in kindergarten are.

There was no way to relate forms filled out by teachers to children's scores except in the case of one-class centers. Therefore, we used for each item the average rating of all staff members of a CDC who submitted a form[1] as the index of the center's belief about its children. Items were scaled from 1 to 5, with 1 representing "much worse." (We recognized the problems inherent in this whole procedure, but we believed that it was worth trying to exploit any information available.)

The center average unadjusted raw score means of each test were then correlated with the center average for each item. The product-moment coefficients are shown in Exhibit 55. The correlations are based on a sample size of 72.

The results of this admittedly gross procedure are interesting, but, like many correlational studies, difficult to interpret. First, the items themselves are highly intercorrelated (except for "Success in Kindergarten," the intercorrelations are all over 0.90). Consequently, the questions of interest are:

- Are the correlations between items for a given test really different?
- Are the correlation coefficients across tests for a given item really different?

The questions can be examined by transforming the coefficients to standard normal variables and calculating the standard error of the

---

[1] A respondent was required to fill out only one form, so the items are impressions about the CDC children as a whole, presumably.

EXHIBIT 55 - CORRELATIONS OF CENTER OPINIONS AND TESTS

Item	PPVT	PSI	VSMS	BI
1. Getting Along	0.289	0.226	0.077	0.382
2. Self-Confidence	0.344	0.290	0.086	0.309
3. Speaking Ability	0.339	0.275	0.116	0.348
4. Manners	0.416	0.331	0.110	0.359
5. Finishing Tasks	0.327	0.267	0.112	0.304
6. Obedience	0.378	0.295	0.138	0.379
7. New Interests	0.313	0.246	0.088	0.353
8. Self-Reliance	0.299	0.251	0.060	0.335
9. Kindergarten	0.305	0.266	-0.039	0.057

difference between two correlations. Differences on the order of 0.30 to 0.35 are significant by this procedure. Thus, it is clear that, with one possible exception for the kindergarten variable, the variations of correlations between items for a given test are not significant. Furthermore, the differences in correlations for given items across tests are not significant either. Thus, this opportunity for this form of assessment of validity did not yield information.[1]

Another aspect of the validity of the tests is the question of the interaction of test and/or test situation and child. There is growing concern that tests such as the PPVT, for instance, may inhibit children by the nature of the situation that they present. Thus, Cline, et al,[2] suggest that the forced choice, receptive language mode of the PPVT is anxiety-provoking to children who see being controlled (forced choice, receptive language) as particularly threatening. Cline suggests that low-income Negro children are especially likely to show withdrawal and/or random behavior under those conditions, and thus to show a performance deficit.[3] Therefore, a test such as the PPVT may not be sensitive to genuine cognitive changes by virtue of its inhibiting effect.

We have no rigorous data bearing directly on the test anxiety hypothesis. The negative correlations of the PPVT with race (Exhibits 49 and 51) are always larger than the negative correlations of the PSI total score and subtests scores with race. This is consistent with the hypothesis that the PPVT, by controlling and thus raising the anxiety level more than the more expressive language PSI (relatively), penalizes particularly those children sensitive to such restriction. However, there are other interpretations possible.

---

[1] One reason for presenting this analysis, gross though it was, was to illustrate some problems that arise in attempting to use different types and sources of data for evaluational purposes.

[2] Cline, M.G., Judith Marshall, and Eunice Stansbury, "Stanford-Binet, PPVT, and Low Income Preschoolers: New Pitfalls for Old Tests." Paper presented to Eastern Psychological Association, April 1966. (A mimeographed copy is available from Dr. Marvin G. Cline, Institute for Youth Studies, Howard University, Washington, D.C., 20001).

[3] Loc. cit.

One set of evidence comes from ratings of test behavior made by testers directly following administration of the PPVT (generally made while the child was doing the DAP). The front of the PPVT test record provides nine descriptors of test behavior arranged in the form of 3-category polar scales. Testers were not provided with criteria for using these scales, nor was the use of them stressed during tester training. Consequently, any consideration of the results of these ratings must be made under strong caveats concerning their reliability and validity. In this same vein, the further confounding factor of observation time should also be noted; by the time that the tester rated the child, he had observed him reacting to the PSI as well as to the PPVT.

With all of these qualifications, we show two types of classifications of the results of the ratings. We selected five variables that appeared to be (1) related to the test anxiety hypothesis, and (2) somewhat conducive to the development of a meaningful discrimination on the part of the tester. Sorting on the basis of variables to be described, we made a count of the number of children rated at the extreme negative end of each variable or item and the remainder of the children. Thus, for each test behavior variable or item, we determined the frequencies of extreme negative ratings (N_L) and non-extreme negative ratings by:

- Race and sex
- PPVT IQ above or below the mean IQ for Group 1 (IQ = 85).

The percentages of children rated at the extreme negative end of an item scale, by race and sex, for S and L,[1] and the two samples combined are shown in Exhibit 56. Percentages of children above and below the Group 1 average IQ who were rated extremely negative, for essentially the same S and L samples and for the M sample as well, are shown in Exhibit 57. The five variables, with the extreme negative descriptor in each case, are:

---

[1] Samples include non-eligible as well as eligible children.

EXHIBIT 56 - PERCENTAGES OF CHILDREN RATED NEGATIVELY ON SELECTED TEST BEHAVIOR VARIABLES

Item	Treatment Level	N	Percent of Total Sample Rated Low on Item (N_L)	N_L	Percent of Low Sample			
					White		Non-White	
					Male	Female	Male	Female
1. Guessing	Short	303	26.1	79	19.0	11.4	35.4	34.2
	Long	278	17.3	48	12.5	12.5	33.3	41.7
	Total	681	18.6	127	16.5	11.8	34.6	37.1
2. Speed of Response	Small	315	11.4	36	11.1	16.7	47.2	25.0
	Long	282	7.8	22	13.6	9.1	40.9	36.4
	Total	597	9.7	58	12.1	13.8	44.8	29.3
3. Verbalization	Short	312	18.3	57	15.8	21.1	38.6	24.5
	Long	281	12.1	34	11.8	8.8	47.1	32.3
	Total	593	15.3	91	14.3	16.5	41.8	27.4
4. Attention Span	Short	313	16.0	50	16.0	18.0	32.0	34.0
	Long	284	20.4	58	8.6	3.4	55.2	32.8
	Total	597	18.1	108	12.0	10.2	44.4	33.4
5. Need for Praise	Short	311	24.1	75	18.7	26.7	37.3	17.3
	Long	278	15.8	44	15.9	6.8	40.9	36.4
	Total	589	20.2	119	17.6	19.3	38.7	24.4

EXHIBIT 57 - ASSOCIATION OF LOW RATINGS ON SELECTED PPVT TEST BEHAVIOR SCALES WITH PPVT IQ

Item	Treatment Level	N	Percent of Total Sample Rated Low on Item (N_L)	N_L	Percent of Low Sample with PPVT IQ <85	≥85
1. Guessing	Short	308	25.6	79	53.2	46.8
	Medium	330	17.9	59	37.3	62.7
	Long	287	16.7	48	54.2	45.8
	Total	925	20.1	186	48.4	51.6
2. Speed of Response	Short	321	11.8	38	78.9	21.1
	Medium	334	9.3	31	74.2	25.8
	Long	294	7.8	23	78.3	21.7
	Total	949	9.7	92	77.2	22.8
3. Verbalization	Short	317	18.0	57	70.2	29.8
	Medium	334	15.3	51	62.7	37.3
	Long	291	11.3	33	75.8	24.2
	Total	942	15.0	141	68.8	31.2
4. Attention Span	Short	317	16.1	51	68.6	31.4
	Medium	334	17.1	57	59.6	40.4
	Long	293	20.1	59	59.3	40.7
	Total	944	17.7	167	62.3	37.7
5. Need for Praise	Short	312	23.7	74	77.0	23.0
	Medium	330	18.2	60	63.3	36.7
	Long	292	15.4	45	62.2	37.8
	Total	934	19.2	179	68.7	31.3

- Guessing: resisted guessing.
- Speed of response: slow.
- Verbalization: taciturn.
- Attention span: distractable.
- Need for praise: much needed.

The reader should bear in mind the varying compositions of the S, M, and L samples in considering these data. (They are unweighted percentages.) He should also bear in mind that there is no indication of the extent to which the extreme negative ratings tended to be the products of only a few testers. It is clear that responses tended to be extremely negative quite a bit less than one-third of the time.

It appears that the Negro children, disproportionately often (see Exhibit 2 for the total sample composition), were rated, following the administration of the PPVT, as having forms of behavior consistent with the test anxiety hypothesis. Further, these extremes of behavior tended to be related to lower standing on the PPVT IQ scale (used here for convenience to control for age). There was also some tendency for the proportion of children rated extremely negative or low in the total treatment level samples to decrease from S to L. The exception occurred in the "attention span" area, and here it is important to bear in mind that L program children were generally younger than S or M types. There is also an indication in Exhibit 56 that, with the exception of "guessing," Negro boys were more frequently rated low on the test behavior variables than Negro girls.

Thus, there appears to be some marginally interpretable evidence consistent with the test anxiety hypothesis, and, indeed, consistent with the hypothesis that test anxiety is higher among Negro than white children, when other things are equal

If it is true that testing, and especially receptive language testing, is particularly threatening to low-income Negro children, then it may be asked whether Head Start programs help such children change their perception and reactions to a situation that is analogous in many ways to the structure of public school classroom operations. The data shown here suggest some tendency for a change in a positive direction, but a change occurring more with white than non-white children.

There is a further, more remote, consideration bearing on validity in some of the data available in this study. Examination of individual items in the PSI, for example, reveals a number of interesting points (see Exhibit 35). The difficulty of most items clearly varies with age, although for some it does not. An example of a non-age sensitive item is Subtest 1, item 4 ("When is your birthday?"), which was quite uniformly difficult for all except two of the three L 6-year-olds. Similarly, item 22 in the same subtest ("Put one car in the middle-sized box") was not much easier for older than for younger children. The question of interest is: are these items difficult for children in the 3- to 6-year-old age range generally, or are they difficult for low-income children? This question cannot be addressed directly without comparative data from non-low-income children. Some of the uniformly difficult items may simply be outside the general vocabulary range or experience of young children. However, in light of the frequently reported lack of sensitivity of culturally deprived children to temporal distinctions, it is interesting to note the more or less uniformly low level of correct identification of times (Subtest 2, items 35, 36, and 37, having to do with identifying the time of year that it is (1) hottest, (2) coldest, and (3) at present). Even for 6-year-olds these were difficult items. It is tempting to take this difficulty as consistent with other findings about the lack of temporal discrimination in poor children.

Were the instruments used appropriate? Or were they too irrelevant, unreliable, insensitive, inhibiting, or overly contaminated by nuisance factors to detect genuine cognitive, social, or behavioral changes associated with Head Start program participation? There is no simple yes or no answer to these questions. We have reviewed evidence available from our own data and offer the following comments as our opinions, based more on impressions than on cold analysis.

1. For purposes of detecting general shifts in performance in a situation calling for use of receptive language skills and/or willingness and ability to operate according to the demands of authority (teacher, tester, etc.)--that is, to play the game--the PPVT seems to be fairly appropriate when differences in raw score means are used as the measures of effects. In our study, the PPVT was generally sensitive to variables to which it ought to be sensitive. To the extent that it simulates one form of situation or relationship which Head Start children, like others, will ineluctably encounter with increasing frequency and seriousness in their public school careers as the system presently works, it matters little whether the changes measured by it are cognitive, motivational, or both. Functionally, the result is the same. The challenge for Head Start is to find and clarify those procedures and techniques that maximize the development of effective cognitive skills, whether such techniques are directed at cognition, motivation, or both. It does not, in general, appear to have done so yet. We are not saying that other tests might not be equally or more appropriate. We are saying that, other things being equal (including, incidentally, administration costs), we think that the PPVT is reasonably appropriate for Head Start program evaluation purposes.

2. We think that the revised PSI is at least as appropriate for evaluation purposes as the PPVT. However, administration costs, including tester training, are higher. The PSI in many respects provides more information that is of operational significance than does the PPVT. Furthermore, the PSI appears a little more sensitive to variables that one would expect it to be, and possibly a little less sensitive to confounding variables. The PSI was more sensitive to the urbanization measure than the PPVT (see Exhibit 45), although we did not examine individual items with respect to that variable. Somewhat more of the total PSI variance was accounted for by the independent variables considered in this study. Thus, the PSI is probably more sensitive to local conditions than the PPVT and as such makes a better instrument for local diagnostic purposes for children who were not extremely

handicapped than does the PPVT. For our purposes, it seemed no more appropriate than the PPVT, and it was substantially more costly to administer and to train testers for than the PPVT. We have reservations about the order in which some items occur on the PSI (see Appendix D). We think that the present grouping of subtest items may enhance whatever test anxiety is inherent in the situation for low-income children. In some ways, the PSI seems constructed more to accommodate the academic standards of test specialists than to provide interpretable information about cognitive content or achievement. Nevertheless, it has at least a face validity for evaluation of Head Start programs that the PPVT lacks.

3. The Behavior Inventory remains an enigma. It told us little about effects (neither, of course, did other instruments). It told us relatively little about sensitivities to background or control variables. A number of teachers found many items ambiguous or hard to answer because of their multidimensionality. We found that the design of the form contributed substantially to the omission of responses by teachers and that a factor analysis of BI subtests gave us meaningless results. We found it difficult to imagine what criteria teachers used in making some ratings. Our overall opinion is that, before the BI is used further for diagnostic or evaluation purposes, systematic investigation and evaluation of it as an instrument should be undertaken. There is some evidence that it is grossly sensitive. However, there are too many uncertainties about what it really is measuring under conditions such as ours to recommend its further general use without more research and evaluation.

4. The Vineland Social Maturity Scale, as employed in this study, seems grossly appropriate, but not worth the cost of tester training and test administration. The negative beta coefficients for center size in the stepwise regression analysis, plus the negative beta for the amount of teacher's pre-school experience, plus the significantly positive age betas (and correlations), suggest a tendency to "report" in terms of an age stereotype or not to report enough information. (Testers

sometimes reported that teachers in larger centers seemed less knowledgeable about or familiar with the children.) The overall average social quotients (SQ's) obtained in this study were about 100, as they should have been if this had been a typical group or if teachers had responded in terms of a typical stereotype. We found evidence of a tester-teacher interaction with the VSMS, and of a tester bias.

It is not clear that the VSMS would be sensitive to effects of Head Start treatment, even if the parent were the respondent. We think that the value of the VSMS in this study was to establish that the children in our sample probably were not conspicuously advanced or retarded for their ages in terms of the skills and abilities examined by the interview.

b. Reliability

The question of reliability is basically the question of error of measurement. It is often assessed by a variety of methods, including observations made at different points in time (e.g., test-retest), or analyses of a single set of observations (e.g., split-half correlations or item statistics analyses).

As stated earlier, there was no opportunity in this study to make any of the standard reliability checks in terms of repeated observations or multiple observers. We did calculate KR-20 coefficients[1] for the PSI subtests, for four age groups, and for each treatment level. The coefficients are presented in Exhibit 58. The coefficients for Subtest 2 are based only on items 27 through 42, since items 43 through 47 were 2-point items. Sample sizes by age and program level are the same as those shown in Exhibit 35.

The coefficients for the 3-year-olds and for the L program's 6-year-olds are based on very few observations. This seems to be reflected in a greater diversity of coefficients for these groups. Otherwise, the coefficients seem much alike. Generally the M level coefficients appear to be higher than those around them.

---

[1] Kuder, G.F. and M.W. Richardson, "The Theory of the Estimation of Test Reliability," Psychometrika, 1937, Vol. 2, No. 3, pp. 151-160.

EXHIBIT 58 - KR-20 COEFFICIENTS FOR PSI SUBTESTS

Subtest	Age	Program Duration		
		Short	Medium	Long
1	3	0.98	0.76	0.72
	4	0.74	0.86	0.73
	5	0.78	0.82	0.77
	6	0.73	0.87	0.57
2(1)	3	0.53	0.62	0.67
	4	0.74	0.72	0.81
	5	0.73	0.80	0.70
	6	0.78	0.82	0.79
3	3	-	0.15	0.60
	4	0.60	0.68	0.63
	5	0.76	0.75	0.75
	6	0.73	0.82	0.61
4	3	0.67	0.73	0.72
	4	0.67	0.78	0.67
	5	0.75	0.80	0.67
	6	0.70	0.72	-

Note: (1) Based on items 27 through 42 only.

By this index the reliability of the PSI subtests appears relatively satisfactory.

For the main part of the study, our concern has been with reductions of total variances. Test reliability or unreliability is simply one factor contributing to an unknown extent to our error variance. We believe that we have specified fairly the reliability of the means, differences, and weights obtained, but there is no way of relating test reliability sums to the sums in those models.

5. Effects of Head Start Which May Not Be Immediately Noticeable

If there are latent effects, by definition we would not have detected them with end testing. There is no question that longitudinal or cohort-tracking studies are vitally needed for major social programs and should, in the long run, provide more reliable and interpretable information than short-term studies such as ours. The dilemma, from a program point of view, is contained in the phrase "in the long run." Short-term studies have the most potential for influencing practical correction of discrepancies or inadequacies before procedures, operations, practices, attitudes, and so on become institutionalized. On the other hand, it appears that results from short-term studies are the more difficult to interpret both on technical and theoretical grounds. We are inclined to think that the trouble with short-term studies is just that; they are simply short-term studies. To the extent that they provide a base for continued observation, their value should be enhanced substantially.

We conclude that there are a number of reasons why we may have failed to detect social or psychological effects of Head Start programs in our samples. We have discussed what seem likely to be the principal ones. For us, the overriding constraint is the lack of information about local program objectives and goals. Without better definitions, it appears that evaluation may continue to be a marginally informative operation.

C. Exceptions

There were two exceptions to the trend of results that warrant comment. One was the significantly low adjusted BI mean for the M treatment level (Exhibit 39). The other was the significant beta coefficient for "weeks" for PPVT and PSI equations for Group 5, children tested in Spanish (Exhibit 45). We shall comment on these in turn.

1. The Behavior Inventory

There are several choices of explanation for the result, depending upon the assumptions accepted. Basically, the result occurred because the model applied essentially the same type of correction to the BI that it did to each of the other measures: it lowered the M mean. In the case of the BI, it did so to compensate for the higher level of every covariate except population, modified by the correlation of the BI with those variables in the overall sample, and by their intercorrelation. It is apparent that the large decrement for the M level BI was associated with race (there were disproportionately more whites in the M sample than in the other two). Decrements beyond that, in diminishing order of magnitude, were contributed by age, poverty, and sex.

If we examine the BI means, adjusted and unadjusted, when age, sex, and race are held constant (Exhibit 43), we find much the same result with two conspicuous exceptions: the M level means are adjusted upwards for the 5-year-old white males and females. The adjusted means show the same V configuration over SML in seven of the 10 groups as the By-Center means,[1] as compared with eight of the 10 for the VSMS, and five of 10 for the PPVT and PSI. Thus, it appears that especially where the two teacher-related instruments are concerned, the V configuration is not just a correction for age, sex, and race inequalities.

In the By-Center regression analysis, the independent variables accounted for about 9 percent of the total BI variance, less than for any of the other measures. In the By-Child stepwise regression analyses, the 13 independent variables used accounted for a little over 8 percent of the variance for Group 3, with age and population having the only significant betas.

---

[1] It will be recalled that none of the F-ratios for the By-Child analyses was significant at the 0.05 level of confidence.

One hypothesis to account for these results is based on the assumptions that the BI provides an accurate index of behavior and that the S, M, and L groups were comparable in behavior by this index at the start of their programs. Given these, it would appear that there is, over time, a shift in the behavior of Head Start children away from forms valued by or acceptable to teachers, followed by a readjustment or trend toward desired forms. Thus, the hypothesis is that Head Start encourages the occurrence of a behavioral change sequence which tends to be (as well as to appear) undesirable or destructive, but which ultimately moves toward positive values. This concept is analogous, in some respects, to the transference sequence expected by many psychotherapists.

A second hypothesis takes account of one further bit of data: the positive correlation of the BI scores with population. This hypothesis is that there is some factor related to the size of community that is related either to the behavior of the children, or to the subjective criteria used by teachers in making the ratings. There could be some form of urban/rural difference in teachers' standard or criteria, or in the behavior of children, or both. PRC has not, however, examined the slopes of the regression on population for S, M, and L to determine whether they were in fact dissimilar. This needs to be done before proceeding with further hypothesizing about factors associated with differences in community size.

A final hypothesis is that we were unlucky. The F-ratio for the BI was significant, but just barely. The exact probability was 0.04. Whatever variables exerted significant control over ratings on the BI were not the ones for which we accounted. If the F-ratio had been just slightly smaller, and thus not significant at (arbitrarily) the 0.05 level, we would have thought no more about it. Uncertainties about the BI have already been discussed in subsection V. B. 4. Until more is known about the properties of the BI as an instrument, further speculation or hypothesizing does not seem worthwhile. In any event, we cannot reasonably conclude that the V configuration of the BI means for S, M, and L samples points to a program or treatment effect.

2. Children Tested in Spanish

The results of the By-Child stepwise regression analyses for Group 5, the children tested in Spanish, contrasted markedly with the results for other groups (see Exhibit 45). Specifically, for the regression equations for the PPVT and the PSI total score, the beta weights for weeks and for hours were significant. Yet more striking, the coefficient for weeks was in both cases negative (the longer the time in program, the lower the score). More striking still was the fact that no other independent variables had significant betas. Finally, the total proportions of variance accounted for were higher for both tests with this group than for any other group (over 55 percent).

Note also in Exhibit 45 that this group had the highest average age (72.67 months) of any group.

On the face of it, the results seem to suggest that the longer (in terms of weeks of program) that Spanish-speaking children stay in Head Start programs, the lower they score on the PPVT and PSI. However, the longer (in terms of hours of daily program) that they stay at a Head Start CDC each day, the higher they score.

It must be noted that the Spanish versions of both tests were not standardized (see Appendix D for protocols and Spanish translations used).

No clear hypotheses emerge from the data. Forty-four percent of the PPVT variance was accounted for by weeks, age, sex, and population, entering the analysis in that order. Hours next accounted for an additional 6 percent, income for the next 3 percent, and family size and center size for the remaining 2 percent. With respect to the PSI, 47 percent of the variance was accounted for by weeks, age, and family size (relatively high negative partial correlation). The next 8 percent was accounted for by mother's education (1-1/2 percent), hours (1-1/2 percent), and income (5 percent). Center size, population, and sex, in that order, account for the remaining 3 percent.

The partial correlations of the independent variables with the two tests are notably different, in some respects. They are shown in Exhibit 59.

EXHIBIT 59 - PARTIAL CORRELATIONS FOR SPANISH-SPEAKING GROUP 5

Independent Variable	Dependent Variable	
	PPVT	PSI
1. Age	0.18	0.16
2. Sex	-0.33	0.15
3. Weeks	-0.37	-0.37
4. Hours	0.39	0.37
5. Mother's Education	0.00	-0.23
6. Population	0.32	0.25
7. Center Size	0.10	0.28
8. Family Income	0.26	0.24
9. Family Size	0.15	-0.30

It is easy to suspect a selection factor operating across the S, M, and L Spanish-speaking samples. However, there is no way of demonstrating it. Furthermore, the partial correlation coefficients, along with the variables accounting for the variance of the PPVT and of the PSI, suggest two quite different mechanisms operating with the two tests.

Whatever the explanatory hypotheses, it is clear that this sample of Spanish-speaking children is substantially behind the other samples on the two individual tests, and on the VSMS as well. (This is true despite the fact that they were presumably given an advantage in language for testing purposes which they do not enjoy in school.)

VI. CONCLUSIONS

A large amount of data has been presented and many questions raised and discussed. Yet further data and discussions are furnished in appendixes to this report. This section will present the primary conclusions that stand out in the context of the overall study. No attempt will be made to list all subsidiary conclusions.

1. Subject to limitations in interpretation imposed by the design of the study, there was no statistically reliable evidence of a change in performance or rating of children in the major eligible samples on four test instruments which could be related to the length of a Head Start program or to the length of time that a child had attended a Head Start center. The conclusion also holds for various subsamples of children of similar age, sex, and race. However, the conclusion loses operational significance for subgroups as sample sizes decrease, since only very large changes can be assessed reliably when samples are very small.

The one significant variation in the test means between duration levels occurred with the Behavior Inventory. Various interpretations or explanations of the deviation were considered. PRC concluded that the variation was probably related to error of measurement and not to effect of the programs.

2. This conclusion does not vitiate the following hypotheses or possibilities concerning 1966 full-year Head Start programs:

- Children improved in many ways not measured by the tests, including health, nutritional status, and attitude toward schools and teachers.
- Children improved measurably on the dimensions measured by the tests or assessed by the rating or interview scales relatively early in their participation in programs.
- Parents, teachers, other staff members, and community organizations benefited from participation or involvement in the 1966 full-year programs.
- Beneficial effects of participation are latent and will become manifest after the children enter school.

None of these possibilities could be examined within the context of this study.

3. There were a number of factors which could have acted to obscure the observation and measurement of a Head Start treatment effect. Of these, the more significant methodologically appear to be:

- Lack of direct evidence that major experimental samples were comparable at the start of the programs.
- Some uncertainty concerning the validity and reliability of at least one of the measuring instruments.
- Lack of specificity of information about needs and goals associated with different programs.

The first is by far the most serious. As a result of the criteria used to identify the programs in the three duration levels studied, the distribution of centers was quite unlike any usual geographic distribution. Whether there were underlying selective factors differentially associated with the emergence of funded programs at different points in time during fiscal year 1966 is a matter of speculation.

4. The concept of an effect is complex and deserves close attention in the evaluation of large-scale programs aimed at changing behavior.

5. There is some evidence that the generally lower performance of Negro children relative to white children, especially as measured by the PPVT raw score, may be the result of a motivational rather than (or as well as) a cognitive factor. The situation that may cause the depressed scores is analogous to the situation and demands of the school classroom. If this is the case, it is certainly a condition which Head Start programs should be trying to correct.

6. Families of the children in the study were very similar in a number of characteristics in the different program levels. The characteristics of staff members and staff structures were generally similar for the three main samples, although some differences were noted. On a very gross basis of measurement, no significant relationship was observed (with one exception) between test scores and the amount of teachers' experience with pre-schoolers or with children from conditions of poverty.

In summary, this study conducted end-of-program tests of samples of children in 72 1966 full-year Head Start Child Development Centers representing programs of three main lengths or durations. The sample of programs ranged from 6 to 36 weeks in length at time of testing, and the average lengths of the three main samples were 12.4, 19.3, and 27.6 weeks. A sample of children in each center was tested with the Peabody Picture Vocabulary Test (PPVT), a test of general verbal ability, and the Caldwell-Soule Pre-School Inventory (PSI), which is designed to measure performance in several areas of social and cognitive achievement. Teachers were interviewed to provide ratings of the children on the Vineland Scale of Social Maturity (VSMS), and teachers also completed ratings of the children on the Operation Head Start Behavior Inventory (BI). The average test scores of the children in short-term centers (6 to 15 weeks) provide the basis of comparison for the examination of effects of Head Start programs on the children.

Statistical analyses of the results were undertaken in which a number of background variables such as age, sex, race, size of town, etc., were taken into consideration. As noted above, the overall result was that there was no significant indication of a general increase of scores with length of program.

The study did not examine the content or structure of the programs in the sample. Nor was any systematic attempt made to rate or evaluate the quality of the programs, personnel, or operations independently of the test scores. Consequently, PRC does not feel that the results mean either that the programs accomplished nothing, or that many possible short- and long-term benefits to children, parents, and staff members did not occur. There are numerous variables involved in an enterprise as complex as Head Start, and the possible impacts and benefits are manifold. With respect to the functions, processes, or skills of children presumably assessed by the instruments used in this study, it appears that, overall, the gains to be expected with the longer programs exemplified by the 1966 full-year sample studied are small. The challenge, in the continued evaluation of programs such as Head

Start, is to discover yet more precise, and at the same time comprehensive, means of depicting the true nature of the total array of benefits, and to translate such findings in further improvements in program design and operations.

APPENDIX A

SAMPLES

In this appendix the subuniverses and eventual S, M, and L sample centers are listed. Exhibits A-1, A-2, and A-3 list the universes of CDC's for S, M, and L samples. The 173 centers are listed by grant and center number, local address, and city and state. In addition, for each program the early estimates of number of classes and of the length of the program are given. Finally, the cumulative weight for each center in the ordering of the subuniverse is listed, as well as the random start for each subuniverse.

Exhibit A-4 lists the S, M, and L centers eventually tested. Centers are identified by grant and center number and city and state.

EXHIBIT A-1 LIST OF ELIGIBLE SHORT-TERM CENSUS SAMPLE CENTERS FROM WHICH THE PRC SUBSAMPLE WAS DRAWN

CDC Grant and Center Number	Local Address	City and State	OEO Region	Early Estimate of Number of		Cumulative Weight[1]
				Weeks of Program	Classes	
6152-001	Hawthorne Elementary School	Mexico, Missouri	6	11	2	2
1036-001	Leominster Child Development Center	Leominster, Massachusetts	1	12	2	4
2124-001	A. Moten Elementary School	Hampton, Virginia	2	12	1	5*[2]
0879-010	D. Zavala Elementary School	McAllen, Texas	5	12	1	6
0739-011	Riverview Housing Project	Kansas City, Missouri	6	12	2	8
0569-003	Riverview Towers	Patterson, New Jersey	1	13	4	12*
5068-002	Caney Public Schools	Caney, Oklahoma	5	13	1	13
0202-005	Whittier School	St. Paul, Minnesota	6	13	1	14
0202-002	Erickson School	St. Paul, Minnesota	6	13	2	16
0512-013	Irving School	Riverside, California	7	13	2	18*
0416-012	Poplar Elementary School	Fontana, California	7	13	2	20
2044-007	Whitesville School	Whitesville, Kentucky	2	14	2	22
5076-004	Talihina Public School	Talihina, Oklahoma	5	14	3	25*
0844-001	Wintergarden Community Action Center	Eagle Pass, Texas	5	14	4	29
5130-001	Western New Mexico University Child Development Center	Silver City, New Mexico	5	14	3	32*
0879-001	Wilson Elementary School	Mission, Texas	5	14	10	37*
5076-003	Le Flore Public School	Le Flore, Oklahoma	5	14	2	39
0739-002	Carver School	Kansas City, Missouri	6	14	2	41
0739-007	Garrison School	Kansas City, Missouri	6	14	2	43*
0739-016	Yeager School	Kansas City, Missouri	6	14	3	46
0554-007	Greenfield School	Montclair, New Jersey	1	15	10	51*
1103-002	Presbyterian Church of Moorestown	Moorestown, New Jersey	1	15	2	53
5178-001	St. Patrick's Education Building	Waelder, Texas	5	15	4	57*
5070-008	Bowlegs School	Bowlegs, Oklahoma	5	15	1	58
5005-020	Travis School	El Paso, Texas	5	15	1	59
5005-021	Villas School	El Paso, Texas	5	15	2	61
5005-009	Franklin School	El Paso, Texas	5	15	3	64*
5005-006	Clardy School	El Paso, Texas	5	15	5	69*

EXHIBIT A-1 (Continued)

CDC Grant and Center Number	Local Address	City and State	OEO Region	Early Estimate of Number of		Cumulative Weight[1]
				Weeks of Program	Classes	
5005-008	Douglass School	El Paso, Texas	5	15	2	71
5005-023	Zavala School	El Paso, Texas	5	15	5	76*
5070-003	Strother School	Seminole, Oklahoma	5	15	1	77
6149-002	Alamosa County Office Building	Alamosa, Colorado	6	15	1	78
7067-144	Phillips Temple C.M.E. Church	Los Angeles, California	7	15	3	81*
7067-137	Church of Christ	Los Angeles, California	7	15	8	86
7067-124	Scott Methodist	Pasadena, California	7	15	3	89*
7067-146	Emmanuel Missionary Baptist Church	Los Angeles, California	7	15	1	90
7067-010	St. Leo's Hall	Los Angeles, California	7	15	1	91
7067-123	Seventh Day Adventist Church	Santa Monica, California	7	15	1	92
7067-092	Cristo Rey Chapel	Los Angeles, California	7	15	2	94
7067-139	St. Paul's Methodist Church	Los Angeles, California	7	15	1	95
7067-151	Holy Faith Baptist Church	Los Angeles, California	7	15	1	96*
7067-129	Cabrillo Recreation Center	Long Beach, California	7	15	2	98
7067-005	Ascension Hall	Los Angeles, California	7	15	1	99
7067-152	True Faith Baptist Church	Compton, California	7	15	2	101
7067-134	Church of Christ	Los Angeles, California	7	15	2	103*
7067-156	God Bibleway Church	Compton, California	7	15	4	107
7067-161	Immanuel Baptist Church	Los Angeles, California	7	15	2	109*
7067-163	Trinity Baptist Church	Los Angeles, California	7	15	3	112
7067-168	Lincoln Memorial Congregational Church	Los Angeles, California	7	15	2	114
7067-177	Bassett School	La Puente, California	7	15	1	115*
7067-179	Keenan School	La Puente, California	7	15	2	117
7067-185	Rosecrans School	Compton, California	7	15	2	119
7067-186	Willard School	Compton, California	7	15	1	120
7067-192	Bursch School	Compton, California	7	15	2	122*
7067-195	Pico School	Pico Rivera, California	7	15	2	124
7067-202	Emerson School	South San Gabriel, California	7	15	1	125

EXHIBIT A-1 (Continued)

CDC Grant and Center Number	Local Address	City and State	OEO Region	Early Estimate of Number of		Cumulative Weight(1)
				Weeks of Program	Classes	
7067-206	Willard School	South San Gabriel, California	7	15	1	126
7067-213	Lee School	Long Beach, California	7	15	1	127
7067-214	Lincoln School	Long Beach, California	7	15	2	129*
7067-221	Winter Gardens School	Los Angeles, California	7	15	2	131
7067-224	Monte Vista School	El Monte, California	7	15	1	132
7067-229	Grayland Avenue School	Norwalk, California	7	15	1	133
7067-231	North San Antonio School	Pomona, California	7	15	1	134*
7067-233	Lexington School	Pomona, California	7	15	2	136
7067-240	Rorimer School	La Puente, California	7	15	1	137
7067-243	Anderson School	Compton, California	7	15	6	142*
7067-244	Carver School	Los Angeles, California	7	15	4	146*
7067-246	Lincoln School	Los Angeles, California	7	15	8	151
7067-248	Washington Children's Center	Pasadena, California	7	15	2	153*
7067-254	Oakwood Playground	Venice, California	7	15	2	155

Notes: (1) The random start was 4.8.
(2) An asterisk denotes that the center was in the original selection of centers to be included in the PRC subsample.

EXHIBIT A-2 LIST OF ELIGIBLE MEDIUM-TERM CENSUS SAMPLE CENTERS FROM WHICH THE PRC SUBSAMPLE WAS DRAWN

CDC Grant and Center Number	Local Address	City and State	OEO Region	Early Estimate of Number of		Cumulative Weight[1]
				Weeks of Program	Classes	
0846-001	Cambridge Neighborhood House	Cambridge, Massachusetts	1	19	4	4*[2]
0846-002	Cambridge Community Center	Cambridge, Massachusetts	1	19	2	6*
0846-004	Thorndike School	Cambridge, Massachusetts	1	19	3	9
0869-015	Neighborhood House Association	Buffalo, New York	1	19	2	11*
1026-002	Columbus Elementary School	South Norwalk, Connecticut	1	19	1	12
2198-001	Williamstown School	Williamstown, Kentucky	2	19	1	13
2084-008	Julius West Junior High School	Rockville, Maryland	2	19	1	14*
2084-004	Broadacres Elementary School	Silver Spring, Maryland	2	19	1	15
4117-001	Bangor-Mount Vernon School	Bay City, Michigan	4	19	1	16
0201-001	Washington Park School	Cincinnati, Ohio	4	19	2	18*
6122-001	Mancos School District	Mancos, Colorado	6	19	2	20
0810-003	Avondale Elementary School	Avondale, Colorado	6	19	2	22
6059-001	Upper Valley Center	Hamilton, Montana	6	19	1	23*
6127-001	St. James R-L School District	St. James, Missouri	6	19	2	25
7010-001	Washougal Head Start Center	Washougal, Washington	7	19	1	26
7010-005	Ridgefield Head Start Center	Ridgefield, Washington	7	19	1	27*
2057-007	Pennington Gap	Pennington Gap, Virginia	2	20	2	29
2057-003	Ely Dale	Ewing, Virginia	2	20	1	30
2057-008	Robins Chapel School	Keokee, Virginia	2	20	1	31*
0526-001	Peabody Heights Head Start Center	Eastman, Georgia	3	20	4	35*
0751-003	Redeemer Lutheran Church	Joliet, Illinois	4	20	2	37
0751-007	Grace Evangelical Lutheran Church	Wilmington, Illinois	4	20	2	39*
5179-001	Farmington Head Start Center	Farmington, Arkansas	5	20	3	42
6140-001	Havelock School	Lincoln, Nebraska	6	20	2	44*
6105-005	Spanish Mission	Colorado Springs, Colorado	6	20	1	45
6105-004	Broadmoor Community Church	Colorado Springs, Colorado	6	20	3	48*
6140-005	Clinton School	Lincoln, Nebraska	6	20	2	50*
6047-001	Kirksville Head Start Center	Kirksville, Missouri	6	20	5	55*

EXHIBIT A-2 (Continued)

CDC Grant and Center Number	Local Address	City and State	OEO Region	Early Estimate of Number of Weeks of Program	Early Estimate of Number of Classes	Cumulative Weight[1]
0483-142	Barrio Las Palmas	Juana Diaz, Puerto Rico	1	21	2	57
1020-002	First Baptist Church	Pittsfield, Massachusetts	1	21	1	58
1020-004	Trinity Methodist Church	Pittsfield, Massachusetts	1	21	1	59*
0115-001	St. Martin De Porres Pre-Kindergarten Center	New Haven, Connecticut	1	21	2	61
0483-120	Bo Capitanejo	Ponce, Puerto Rico	1	21	1	62
0380-008	Faith Presbyterian Church	Washington, D. C.	2	21	2	64*
0783-007	Harrisburg Elementary School	Harrisburg, Ohio	4	21	1	65
0783-004	Finland Elementary School	Columbus, Ohio	4	21	2	67
6157-001	Cedar Creek School	Cedar Creek, Missouri	6	21	1	68*
6157-006	Hurley School	Hurley, Missouri	6	21	1	69
6003-004	Derby Elementary School	Commerce City, Colorado	6	21	2	71*
2059-060	Evangelical United Brethren Church	Philadelphia, Pennsylvania	2	22	3	74
2059-069	Christ Tabernacle Baptist Church	Philadelphia, Pennsylvania	2	22	2	76*
2059-073	Tindley Temple Methodist Church	Philadelphia, Pennsylvania	2	22	2	78
2059-078	Russell Tabernacle CME Church	Philadelphia, Pennsylvania	2	22	2	80*
0202-013	Lincoln School	St. Paul, Minnesota	4	22	2	82
5225-005	Jones Center	Abilene, Texas	5	22	1	83*
5225-008	Valley View Center	Abilene, Texas	5	22	3	86
0143-001	Denver Gospel Hall	Denver, Colorado	6	22	4	90*
0143-029	Elizabeth	Denver, Colorado	6	22	1	91
0143-030	Northside Community Center	Denver, Colorado	6	22	2	93*
0143-034	Our Lady of Grace	Denver, Colorado	6	22	2	95
7044-004	Tuolumne Elementary School	Modesto, California	7	22	2	97*
2044-003	Western School	Owensboro, Kentucky	2	23	2	99

Notes: (1) The random start was 2.2.

(2) An asterisk denotes that the center was in the original selection of centers to be included in the PRC subsample.

EXHIBIT A-3 LIST OF ELIGIBLE LONG-TERM CENSUS SAMPLE CENTERS FROM WHICH THE PRC SUBSAMPLE WAS DRAWN

CDC Grant and Center Number	Local Address	City and State	OEO Region	Early Estimate of Number of: Weeks of Program	Early Estimate of Number of: Classes	Cumulative Weight(1)
0202-012	Jefferson School	St. Paul, Minnesota	4	29	3	3
0869-003	Baker Homes Development	Lackawanna, New York	1	30	4	7*(2)
2059-067	Jones Memorial Baptist Church	Philadelphia, Pennsylvania	2	30	3	10*
2059-062	Holy Trinity Baptist Church	Philadelphia, Pennsylvania	2	30	2	12
0101-027	Parker	Chicago, Illinois	4	30	1	13
0101-018	Hinton	Chicago, Illinois	4	30	3	16*
0473-004	Edwards School	New Orleans, Louisiana	5	30	2	18
0473-006	Guste School	New Orleans, Louisiana	5	30	3	21*
0473-009	Henderson School	New Orleans, Louisiana	5	30	2	23
0473-015	Laurel-McDonogh No. 1	New Orleans, Louisiana	5	30	1	24
0473-026	Ricard School	New Orleans, Louisiana	5	30	2	26
0473-021	McDonogh No. 32	New Orleans, Louisiana	5	30	2	28*
0636-010	Long No. 1	Omaha, Nebraska	6	30	2	30
0636-014	Pershing	East Omaha, Nebraska	6	30	2	32
0636-004	Conestoga No. 2	Omaha, Nebraska	6	30	2	34*
7044-005	Riverbank Pre-School	Riverbank, California	7	31	8	39*
8014-001	Bear Creek	Lantry, South Dakota	6	32	1	40
8014-005	Green Grass	Eagle Butte, South Dakota	6	32	1	41
8014-009	Thunder Butte	Dupree, South Dakota	6	32	1	42
0869-007	School No. 32	Buffalo, New York	1	33	2	44*
0869-008	School No. 33	Buffalo, New York	1	33	3	47
0101-001	Gladstone	Chicago, Illinois	4	33	3	50*
0101-005	Burroughs	Chicago, Illinois	4	33	2	52
0101-011	Field	Chicago, Illinois	4	33	2	54
0101-019	Jackson	Chicago, Illinois	4	33	2	56*
0101-024	Medill	Chicago, Illinois	4	33	5	61*
0101-026	Ogden	Chicago, Illinois	4	33	2	63

EXHIBIT A-3 (Continued)

CDC Grant and Center Number	Local Address	City and State	OEO Region	Early Estimate of Number of		Cumulative Weight(1)
				Weeks of Program	Classes	
0101-030	Skinner	Chicago, Illinois	4	33	3(3)	64(3)
0101-032	Thorp, J. N.	Chicago, Illinois	4	33	4	68*
0101-038	St. Ambrose	Chicago, Illinois	4	33	2	70
0101-042	St. Cecilia	Chicago, Illinois	4	33	2	72*
0101-047	St. Mel	Chicago, Illinois	4	33	2	74
0101-051	Green Homes Branch of Lower N. C.	Chicago, Illinois	4	33	3	77*
0101-053	Hull House	Chicago, Illinois	4	33	6	82
0101-059	Greater Institutional AME Church	Chicago, Illinois	4	34	2	84*
0101-060	Greater St. John Church	Chicago, Illinois	4	34	4	88*
0463-005	Salvation Army Center	Stamford, Connecticut	1	35	2	90
0479-009	Inkster Pre-School Program	Inkster, Michigan	4	35	2	92
1063-002	Village Fours	Manaroneck, New York	1	36	4	96*
2059-008	Cohocksink Center	Philadelphia, Pennsylvania	2	39	2	98
2059-026	Mount Zion Baptist Church	Philadelphia, Pennsylvania	2	39	4	102*
2059-031	Norris Square United Presbyterian Church	Philadelphia, Pennsylvania	2	39	3	105*
2059-035	St. Jacobus Lutheran Church	Philadelphia, Pennsylvania	2	39	2	107
2059-047	Wilson Park	Philadelphia, Pennsylvania	2	39	2	109
2059-021	Johnson Homes	Philadelphia, Pennsylvania	2	42	2	111*
2059-018	Harrison Plaza	Philadelphia, Pennsylvania	2	42	3	114
2059-049	Y. M. C. A.	Philadelphia, Pennsylvania	2	42	3	117*
2059-014	First African Baptist Church	Philadelphia, Pennsylvania	2	42	5	122*
2059-082	Dunbar School	Philadelphia, Pennsylvania	2	42	2	124
2059-040	St. Stephen's Methodist Church	Philadelphia, Pennsylvania	2	42	5	129*
2059-009	Cornerstone Baptist Church	Philadelphia, Pennsylvania	2	42	9	134*

Notes: (1) The random start was 5.0.
(2) An asterisk denotes that the center was in the original selection of centers to be included in the PRC subsample.
(3) Arithmetic error.

EXHIBIT A-4 LIST OF CENTERS TESTED BY PRC

Short (15 Weeks or Less)		Medium (17 to 23 Weeks)		Long (25 Weeks or More)	
7067-244	Los Angeles, California	6157-001	Cedar Creek, Missouri	2059-047	Philadelphia, Pennsylvania
7067-124	Los Angeles, California	5225-005	Abilene, Texas	0101-018	Chicago, Illinois
7067-177	LaPuente, California	6127-001	St. James, Missouri	0101-032	Chicago, Illinois
7067-243	Compton, California	0751-007	Wilmington, Illinois	0101-042	Chicago, Illinois
7067-231	Pomona, California	0869-015	Buffalo, New York	0101-051	Chicago, Illinois
0879-001	Mission, Texas	0201-001	Cincinnati, Ohio	0101-024	Chicago, Illinois
5076-004	Talihina, Oklahoma	2057-008	Keokee, Virginia	0101-060	Chicago, Illinois
5005-023	El Paso, Texas	0143-030	Kenver, Colorado	1063-002	Mamaroneck, New York
0739-007	Kansas City, Missouri	7044-004	Modesto, California	0101-001	Chicago, Illinois
2124-001	Hampton, Virginia	6003-004	Commerce City, Colorado	0101-059	Chicago, Illinois
5005-006	El Paso, Texas	0526-001	Eastman, Georgia	0473-026	New Orleans, Louisiana
0569-003	Paterson, New Jersey	6140-001	Lincoln, Nebraska	0473-004	New Orleans, Louisiana
5005-009	El Paso, Texas	6047-001	Kirksville, Missouri	0869-003	Lackawanna, New York
0554-007	Montclair, New Jersey	7010-005	Ridgefield, Washington(1)	2059-031	Philadelphia, Pennsylvania
5178-001	Waelder, Texas	6140-005	Lincoln, Nebraska	2059-067	Philadelphia, Pennsylvania
5130-001	Silver City, New Mexico	0101-019	Chicago, Illinois(2)	2059-018	Philadelphia, Pennsylvania
7067-161	Los Angeles, California	0636-004	Omaha, Nebraska(2)	2059-035	Philadelphia, Pennsylvania
7067-163	Los Angeles, California	7044-005	Riverbank, California(2)	2059-062	Philadelphia, Pennsylvania
7067-254	Venice, California	2059-060	Philadelphia, Pennsylvania	2059-0[illegible]6	Philadelphia, Pennsylvania
7067-152	Compton, California	2044-003	Owensboro, Kentucky	2059-049	Philadelphia, Pennsylvania
1103-002	Moorestown, New Jersey	2059-069	Philadelphia, Pennsylvania	2059-040	Philadelphia, Pennsylvania
7067-134	Los Angeles, California	2059-073	Philadelphia, Pennsylvania		
7067-192	Compton, California	0810-003	Avondale, Colorado		
7067-144	Los Angeles, California	6059-001	Hamilton, Montana		
		1020-002	Pittsfield, Massachusetts		
		1020-004	Pittsfield, Massachusetts		
		2059-078	Philadelphia, Pennsylvania		
N = 24		N = 27		N = 21	

Notes: (1) This was in the list of L centers when the PRC sample was drawn.

(2) Program duration was 16 weeks.

APPENDIX B

STATISTICAL MODELS AND ANALYSES

A. Introduction

This appendix states the models underlying PRC's analyses and the reasons why it is felt that they are applicable. Subsection B deals with the analyses of covariance. One of these is far and away the most important of the analyses and an attempt is made to give an idea of its power. The adequacy of linear models is addressed in subsection C (dealing with regression), rather than in subsection B. Subsection D explains some of the peripheral analyses. Cutting across the three sections is the important subject of the center (cluster) effect.

B. The Analyses of Covariance

The analysis most informative about effects of S, M, and L is the By-Center analysis of covariance; indeed, the study was designed for this sort of analysis. Center averages were used as observations and concomitant variables. Because centers were clusters in the sampling, the scores of children within a center are correlated, while center averages are independent. Moreover, least squares methods are relatively insensitive to inequality of variances, so the differing numbers of children in the center averages should do little harm. Thus, in terms of closeness to a useful least squares model, center averages are good.

Two disadvantages of the center average model should be mentioned. Comments about the first, loss of power, are vexed by the question: relative to what? There is a center component in the error sum of squares; if this could be estimated and removed, PRC's test would be more powerful. (It should be noted that the corrections for age, race, etc. probably considerably reduce what might otherwise be taken for a center component.)

The second disadvantage is that it is somewhat harder to give precise meaning to the model's parameters than is desirable because centers were chosen with probability proportional to the number of classes in the center. For example, the overall mean is not that of a center chosen at random but, perhaps more relevantly, an approximation to that of a child chosen at random. There is no doubt, however, about the relevance of the model's parameters, and of the size and direction of their estimates which PRC would have liked to have seen.

The model is 4-variate. Let Y_{ij} be the vector of four average scores for the ith center of type j ($i = 1, \ldots, n_j$; $j = S, M, L$) . Assume that

$$E\left\{Y_{ij}\right\} = \mu_j + \beta\,(X_{ij} - \overline{X})$$

and

$$V\left\{Y_{ij}\right\} = \Sigma$$

where β = a 4 x 5 matrix of coefficients

X_{ij} = the vector of the five concomitant variables (age, sex, race, population, and poverty)

$\overline{X}$ = the overall average of the Xij

This model differs from a composite of four univariate models only in the allowance for non-zero off-diagonal elements of Σ --i.e., for correlation between the test scores.

A ratio of determinants called Wilks' λ yields the likelihood ratio test of the hypothesis

$$H : \mu_S = \mu_M = \mu_L$$

(It should be noted that, since the μ's are 4-vectors, the alternatives to H are complex.) Lambda has a distribution that in this case (two treatment degrees of freedom) can be exactly transformed to an F-distribution.

The estimates of all parameters of the model except those involving the off-diagonal elements of Σ are precisely those of the individual univariate models. As a result, the four F-ratios are available, if somewhat hard to interpret simultaneously. Corresponding to any one of the F-ratios are the Scheffé confidence intervals with a coefficient applying simultaneously to all contrasts.

The confidence intervals for the treatment contrasts give a fair idea of the power of one of the univariate tests. More directly, the power of a univariate analysis of covariance can be approximated by noting that the μ_j bear the same relation to σ^2 times the non-centrality parameter δ^2 that their estimates, the adjusted treatment means, bear to the treatment sum of squares. The treatment sum of squares, which is never larger than the sum of squared deviations of the adjusted treatment means from the overall mean, is 79, 83, and 59 percent of it for the PPVT, PSI, and BI, respectively. As an approximation, then, to avoid the laborious recalculation of the non-centrality parameter for each set of putative μ_j's, we shall assume that $\sigma^2\delta^2$ is always 70 percent of the adjusted treatment mean sum of squares. (It should be noted that the above percentages change partly because the differences between the adjusted means change; in fact, PRC investigated power for only one test, the PPVT).

We take as σ^2 its estimate, 19.2; because the difference between 61 and infinite degrees of freedom in the denominator has small effect on the power charts, the non-central chi-square tables may be used. For the configuration of μ_j's we take $\mu_L = \mu_S + \Delta$ and $\mu_M = \mu_S + \frac{21}{43}\Delta$. The results are shown in Exhibit B-1.

EXHIBIT B-1 POWER OF PPVT F-TEST

Δ	Sum of Squares	Approximate Non-Centrality Parameter	Power of a 5% Test
2	43	1.57	0.19
4	172	6.27	0.60
6	387	14.10	0.93

Thus it can be seen that a 2-point PPVT spread would probably not have been picked up, while a 4-point spread probably would have.

The By-Child analyses of covariance have explicitly the same model as the By-Center analysis above. Of course, β is now 4 x 4, and the four concomitant variables are population, income, family size, and hours per day. Sex, age, and race specify the analysis. Of course, there are correlations between children of the same center, and the effect of ignoring these correlations in the By-Child analyses is simply not known. The alternatives, however, would have been either to destroy (in some other way) the orthogonality of the design or fractionate and discard data to the point where cell sizes were minute and degrees of freedom squandered on hypotheses of no interest. Given the desirability of doing some By-Child analyses, the ones performed seem as good as possible.

Again the confidence intervals give a fair idea of the power of the tests with the largest and smallest numbers of degrees of freedom. However, with respect to interpretability, there is the question of simultaneity. To the extent that the ten λ's are correlated, we have a good idea of how: they are repeated observations of the same thing. Therefore, the assumption that the λ's are independent will err on the side of overestimating significance. Under the independence assumption, the estimated significance levels can be combined by taking minus two times their logarithm and adding. The result, shown in Exhibit B-2, is 29.78, which is the 92nd percent point of the chi-square distribution with 20 degrees of freedom--i.e., significant at the 8 percent level. In view of the correlation, we must lean toward the conclusion that the one significant λ is happenstance. The yet more muddled question of the significance of the 40 F-ratios is obviated by the fact that none of them is significantly large--i.e., at the 5 percent level.

C. Regression

A By-Center multivariate multiple regression was done with exposure in weeks as one of the independent variables. There were

EXHIBIT B-2 COMBINATION OF THE LAMBDA'S

Number of Analysis	Lambda	Total Number of Children in S, M, and L	Significance Level $\hat{\alpha}$	-2 log $\hat{\alpha}$
1	0.2970	18	0.2000	3.22
2	0.0006	12	0.0060	10.24
3	0.5484	33	0.0900	4.82
4	0.9953	87	0.9997	0.00
5	0.6417	32	0.2500	2.78
6	0.9496	68	0.9400	0.12
7	0.7564	50	0.1900	3.32
8	0.6977	42	0.1700	3.54
9	0.8737	55	0.6400	0.88
10	0.8689	51	0.6500	0.86
Total				29.78

two main reasons for doing this. First, it was a way of using in an analysis the data from the children who were "ineligible" for the analysis of covariance. Second, PRC believed that there would be a value in being able to examine the beta coefficients of the independent variables taken together.

If U_i is the 4-vector of average scores from the ith center, the model is

$$E\left\{U_i\right\} = \alpha + \beta X_i$$

and

$$V(U_i) = \Sigma$$

where β = a 4 x 6 matrix of coefficients

X_i = the center's vector of concomitant variables (the same as those of the analysis of covariance, although calculated including the ineligible children, plus exposure.)

Again, the allowance for non-zero off-diagonal elements in Σ is the only difference from the usual four univariate models; the parameter estimates are the same with the addition of estimates of the intertest covariances (correlations) about, of course, the assumed means shown above.

PRC assessed the adequacy of the univariate regression models and, indirectly, the adequacy of the similarly linear analysis of covariance models. The four sets of residuals were obtained. The PPVT residuals were plotted against the fitted values $\hat{Y}_i$ and the reassuring result is shown as Exhibit B-3. Exhibit B-4 details a comparison of the range of each of the four sets of residuals divided by the estimate of the appropriate standard deviation with the 10 percent and 0.5 percent points of the (only approximately relevant) distribution of the ratio of range to standard deviation in a single sample of 63. The PPVT,

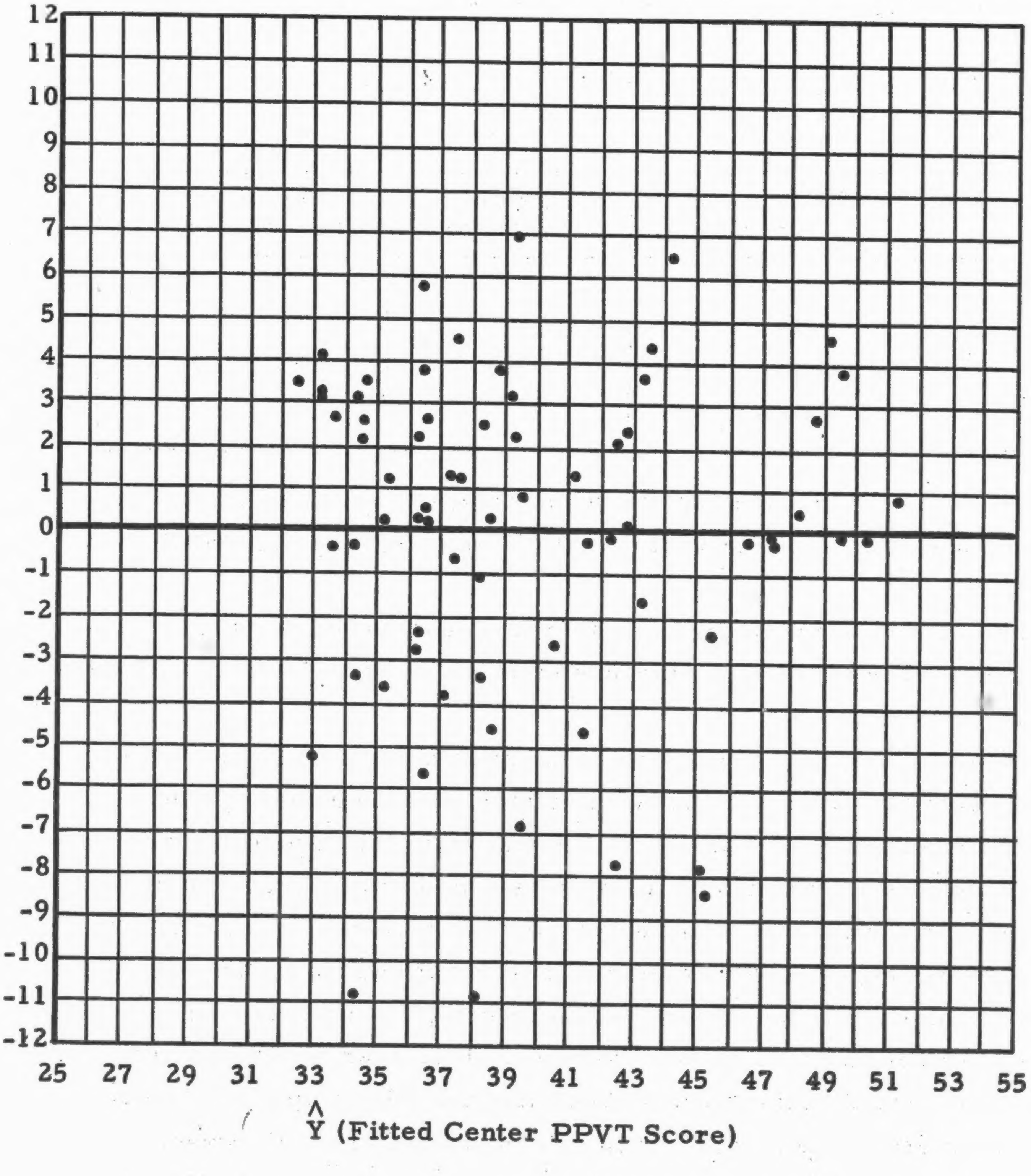


EXHIBIT B-3 PPVT RESIDUALS SCATTER PLOT

PSI, and BI residuals are well within bounds, but there is an outlier among the VSMS residuals (probably only one), for the replacement of the smallest residual by the next smallest would have reduced the normalized range to 3.96 (mutatis non mutandis).

EXHIBIT B-4 RANGE OF RESIDUALS

Test	Normalized Range	Upper 10% Point	Upper 0.5% Point
PPVT	4.42	5.32	6.13
PSI	4.21		
BI	5.33		
VSMS	7.03		

It is unlikely that the results of the analyses would be much changed under a scheme for coping with the one big negative VSMS residual.

The details of the failure of either an age-squared variable or a sex-times-race variable to improve significantly the PPVT fit appear in Section IV.

By-Child stepwise regression analyses were done for each of the four test scores. Since the program printed out all possible product-moment correlations, it was possible to estimate the intertest correlations about the linear mean. There was considerable selection of inputs to even the largest stepwise regression analysis, since only those eligible English-tested children (517 out of 764) who gave the Bureau of the Census information about income, family size, and mother's education could enter. However, their average scores and demographic characteristics look much like the full sample's. The two groups requiring information on father's education are, of course, different from the population as a whole, but not startlingly unlike each other.

The stepwise regression analysis is a series of regressions, each of which includes one more independent variable in addition to those of its predecessor. After the first step, the additional variable chosen is the one that yields the largest studentized coefficient (T-value) in combination with the variables of the previous step. Exposure was always entered first--i.e., the first step was always the simple regression of score on exposure. Once entered, variables could not drop out at a later step. As a result, the effect of entering exposure at any step could be inferred. This procedure did not, of course, yield the combination of independent variables such that the addition of exposure would have produced the largest possible gain in proportion of variance explained. Indeed, short of exhaustion, there is no procedure. From the results, it is hard to believe that the stated combination would not change radically from sample to sample, but from these poor substitutes for a design it is difficult to make an inference of even the slightest delicacy.

D. Other

Means for various subgroups were calculated independently of the above analyses and reported along with a standard error (SEM). This SEM was calculated as for the overall average in a one-way random effects analysis of variance, where the random effect is the center effect. More precisely, for any attribute let X_{ij} be the measure of it for the ith child in center j $(i = 1, \ldots, n_j; j = 1, \ldots, k)$. Let $N = \sum_j n_j$. We assume .

$$X_{ij} = \mu + a_j + e_{ij}$$

where the center effects a_j and the errors e_{ij} are independent among and between themselves. Assume common center and error variances and name them as follows:

$$\sigma_A^2 = \mathrm{Var}\left\{a_j\right\}$$

$$\sigma_e^2 = \mathrm{Var}\left\{e_{ij}\right\}$$

Then, the variance of the mean $\overline{X}$ is given by

$$\mathrm{Var}\left\{\overline{X}\right\} = \frac{\sum_j n_j^2}{N^2}\,\sigma_A^2 + \frac{1}{N}\,\sigma_e^2$$

Since unbiased estimates of σ_A^2 and σ_e^2 are given by

$$\hat{\sigma}_A^2 = \frac{(K-1)(MS_b - MS_w)}{N - \frac{\Sigma n_j^2}{N}}$$

and

$$\hat{\sigma}_e^2 = MS_w$$

where MS_b and MS_w are the between- and within-mean squares, respectively, from the analysis of variance table, a reasonable estimate of $\mathrm{Var}\left\{\overline{X}\right\}$ is obtained by replacing σ_A^2 and σ_e^2 in the statement of the variance of the mean given above by their hatted counterparts. The reported SEM's are the square roots of the estimates of $\mathrm{Var}\left\{\overline{X}\right\}$ so obtained. In those rare cases where the $\hat{\sigma}_A^2$ component was so negative that the estimate of $\mathrm{Var}\left\{\overline{X}\right\}$ would have been negative, the σ_A^2 component was taken as zero.

For each of ten centers we calculated the F-ratios of between-class to within-class mean square errors for the four scores (see Appendix E, subsection C). None of the PPVT ratios, for example, was significant at the 5 percent level. To get some idea of what kind of PPVT differences might have yielded a significant F, it should be noted that, for an 80 percent chance of at least one of ten tests to be significant, each test must have a power of 15 percent. We shall solve for the sort of mean difference needed to give 15 percent power.

Five of the centers had four classes, while five had two classes. Assume that the means of the 4-class centers are equally spaced over an interval Δ_4 and that three scores are observed from each of the four classes. Then, for a power of 15 percent, Δ_4 must be about 1.2 σ. If we assume that seven scores are observed from each of the classes in the 2-class centers, a 5 percent F-test with power 15 percent is obtained when the mean difference Δ_2 is about 0.57 σ.

The average of the ten mean square errors is 114, which suggests taking $\sigma = 10.7$. Thus, if the 4-class PPVT means were spread over 12.8 points within each center and the 2-class PPVT means were 6.1 points apart, we should have had a significant F with a probability of 0.8.

APPENDIX C

TESTERS: SELECTION, TRAINING, AND SUPERVISION

A. Selection

Candidates for Head Start testers were sought from a variety of sources, including colleges and universities in the Washington, D.C., Boston, and Philadelphia areas; the Peace Corps Placement Office; the United States Employment Service Professional Placement Office; and Planning Research Corporation (PRC) personnel. The people who were especially helpful in suggesting candidates were Mr. William Campbell and Mrs. Gladys George (United States Employment Service), Mrs. Leon Sharpe and Mrs. James Russell (Washington Opportunities for Women), Mr. Robert Calvert (Peace Corps), Dr. Philip Porter (Harvard University), and Mrs. Louise Crenshaw (Bryn Mawr College). Dr. Porter actually recruited and interviewed candidates in the Boston area, and Mrs. Crenshaw scheduled interviews at Bryn Mawr College so that a PRC representative could personally interview 27 young women at the college during the course of one day. Mr. Calvert prepared a memorandum and distributed copies to returned Peace Corps volunteers. The Head Start staff at PRC prepared a memorandum describing the necessary qualifications for testers which was distributed to PRC personnel.

Eventually, 213 candidates were personally interviewed for positions as Head Start testers, and 25 were selected to undergo the intensive tester training programs. Of the 25 who participated in the two training programs, 19 actually tested Head Start children; 3 of the 25 were "selected out" after the training program; and the remaining 3 were held in reserve as substitute testers, but there was no opportunity for them to test.

The following qualifications were considered to be of prime importance in the selection of the testers:

- A college degree in an area related to education, psychology, sociology, or guidance;
- Experience with disadvantaged children;
- Language fluency in Spanish as well as English;
- Flexible summer schedule, with no travel restrictions;
- Successful completion of the tester training program conducted by Planning Research Corporation;
- Favorable recommendation by a college faculty member.

Exhibit C-1 (page C-7) lists the testers who actually tested and presents some descriptive data on each. It should be noted that all testers submitted recommendations from faculty members, participated in an intensive training program (described below), and were supervised periodically during their testing in the various Head Start Child Development Centers.

B. Training

Two training programs were conducted, because some of the testers who were completing the school year were not available for a mid-May program. However, since the testing of children in six centers was scheduled to begin on May 16, at least six testers had to be available and trained by that date. The first tester training program was held May 12-15, 1966, at the Center of Adult Education, University of Maryland. Fifteen tester-trainees attended, including two permanent PRC personnel who were trained as a reserve or emergency test corps. The second training program, also at the Center of Adult Education, was held June 8-11, with twelve participants.

Training was provided not only in the proper administration of the various tests (of which the Vineland Social Maturity Scale interview was by far the most demanding), but also in sampling procedures, adult-adult relationships, test scoring, and general operational procedures to be followed. The scope and intensiveness of the training programs may be seen in Exhibits C-2 and C-3 (pages C-8 through C-12), which list the schedules followed in the two programs.

In addition to the formal training sessions, testers were provided with a handbook containing detailed statements of procedures and other instructions for mnemonic purposes. The sections on test administration and test scoring are reproduced in Exhibit C-4 (pages C-13 through C-19) to give the reader a detailed account of PRC's approach to testing and to make clear the rules by which the various tests were scored.

C. Supervision

Each tester was observed at least once during his testing. Testers who tested at many centers were, of course, observed more frequently. In fact, one or two PRC supervisors were in the field during all weeks of testing. In addition, PRC maintained a 24-hour telephone answering service expressly for testers, so that any problem could be handled by the PRC staff in Washington at any time. The testers were observed as they tested children and interviewed teachers. Usually, the supervisor sat behind the child or interviewee to minimize interference with the testing. Supervisors took notes during test sessions but, of course, never interrupted a session with a comment of any kind. In a private conference following the testing and interviewing, the supervisor discussed with the tester any areas needing improvement. Each supervisor left a memorandum of each conference with the tester, as a reminder.

Some of the kinds of statements, suggestions, and criticisms made by supervisors to testers in the field about their work are presented below.

Behavior Inventory

1. Examine carefully to see that every line has one check only.
2. Fill in the current week of operation of the center.
3. Fill in under "examiner" the name of the person who is filling out the BI--not the tester's name.
4. Put your own tester ID number in upper lefthand corner.
5. Write out child's name in space provided.
6. Put total BI score in space marked "school."
7. There should be 25 positive scores and 25 negative scores.

8. Fill in each child's 11-digit identification number in spaces provided.
9. Use the new cut-out score sheets to score and check.

Preschool Inventory

1. Use the PSI manual to ask every question. Remember "OVERCONFIDENCE KILLS." Do not deviate in any way from the procedures described and explained in the manual.
2. In the lines provided next to questions 43-47, fill in the substance of the child's answer.
3. Once the child has made one answer to a question, do not probe, no matter how absurd the answer may be. Do not prompt child to make correct response.
4. Unless previously specified in PSI manual, do not ask the questions more than once; simply say "do as I said" if the child seems in doubt as to what he should do.
5. Don't "rush" a child through the test. Some children just need time to think, without probing or pushing. However, do not wait too long a time if the child becomes restless; he may be stalling or may not wish to answer the question.
6. Do not repeat what the child said; you will tend to reinforce the response.
7. Be sure to use random and intermittent responses of "good," "fine," etc.
8 Use phrasing as dictated by PSI manual.

Peabody Picture Vocabulary Test

1. Fill out the front side of the PPVT record blank as soon as you finish testing the child. A good time to do this is while you are waiting for the child to complete the DAP.
2. Speak clearly and loudly enough so the child has every chance to hear the correct word.
3. Start at about item 15 unless the child seems unusually slow on the PSI. Remember that a child gets easily bored if he is able to score in the 50's, so give him a chance by starting

with item 10-15, rather than 1. You can always go back to earlier items, as per the PPVT manual.

4. Memorize the instructions from the PPVT manual so you can say them in a conversational tone.
5. Don't count errors with your pencil. Use your eyes.
6. Score each PPVT question as you ask it; do not ask a group of questions and then score, as this induces another level of error.
7. Vary your phases as the PPVT manual suggests, so that you are not always saying "point to...."

Draw-A-Person

1. Give the child a pencil with no eraser--children this age love to draw something and then erase it.
2. Don't ask, "Is that a person?" or "Would you like to add anything else?"
3. Do something else (e.g., complete records on the PPVT) while the child is drawing. It may make him nervous if you seem to be waiting for him to complete his drawing.

Vineland Social Maturity Scale

1. Use phrases that tend to avoid putting words in interviewee's mouth:
    a. Tell me how (child) does...
    b. Describe how...
    c. Does (child) do...
    d. Do you think (child) is able to...
    e. Tell me more about how (child) does...
2. Never ask a direct question using key phrases from the Vineland manual.
3. Discuss the questions in the order described in the manual, beginning with Self Help-General and ending with Socialization.
4. Let teacher "talk herself out" if she once gets started, even if she deviates from the original question. Her statements will provide background information that will give you ample opportunity for asking later questions.

5. If you have to ask the same teacher about several children, vary your questioning techniques and phrases so that she is forced to think about each child individually. Teachers sometimes say, in response to a question, "they all do that." Encourage her to discuss children as individuals.

EXHIBIT C-1 HEAD START TESTERS

ID No.	Name	Age	Sex	Education	Recruited Through	Special Qualifications
01	Paula Alexander	21	F	CG(1)	Bryn Mawr College	Fluent in Spanish; previous testing experience
02	Marilyn Ambrose	24	F	CG	Peace Corps	Fluent in Spanish
03	Lorinda Lou Beller	23	F	CG	PRC	Fluent in Spanish; previous testing experience; education major
04	Bernice Borak	22	F	CG	PRC-American University	Fluent in Spanish; community child care experience
05	Louis Carrillo	23	M	CG	PRC-American University	Fluent in Spanish; linguistics major
06	Joan Cavallaro	21	F	S(2)	Bryn Mawr College	Previous testing experience; anthropology major
07	Barbara Daum	19	F	S	Georgetown University	Nursery school experience; sociology major
08	Susan Farrington	27	F	CG	Peace Corps	Fluent in Spanish; teacher
09	Thomas Grave	23	M	CG	George Washington University	Camp counseling, young children; psychology major, college; psychology graduate student
10	Mary Killeen	24	F	CG	Peace Corps	Fluent in Spanish
11	Carolyn Lawall	23	F	CG	Dr. Philip Porter (Boston)	Previous testing experience; psychology major
12	Louise Le Bourhis	27	F	CG	American University	Teacher; previous testing experience
13	David Lewis	24	M	CG	Peace Corps	Previous testing experience; psychology major
14	Patti Lowery	23	F	S	PRC	Previous testing experience; psychology major
15	Margaret Mattis	22	F	CG	PRC-Catholic University	Teacher; previous testing experience; educational psychology graduate student
16	Carl Morgan	19	M	S	PRC	Extensive experience in community youth and recreation work
17	Robert O'Callaghan	22	M	S	PRC	Community work with young children
18	Ruth Rodisch	20	F	S	Bryn Mawr College	Previous testing experience; tutoring
19	Jeanne Schpok	29	F	CG	PRC	Kindergarten teacher; pediatric nurse; psychology graduate student

Notes: (1) College graduate
(2) College student

EXHIBIT C-2 SCHEDULE FOR FIRST HEAD START TESTER TRAINING PROGRAM, MAY 12-15, 1966

Thursday, May 12	
9:00 p.m.	Check-in at Center of Adult Education, University of Maryland, Adelphi Road and University Boulevard, College Park, Maryland (phone: 301-779-5100)
	Free time; assembling of test kits
Friday, May 13	
8:00-9:00 a.m.	Breakfast
9:00-9:30 a.m.	Check-in for local trainees
9:30-10:30 a.m.	General orientation and background Dr. Allen Ferguson: Orientation to PRC Dr. H. Russell Cort: Orientation to Head Start Project
10:30-11:45 a.m.	Mr. Charles McDaris: Project administration
11:45-12:45 p.m.	Lunch
12:45-1:45 p.m.	Dr. R. Ann O'Keefe: PPVT - introduction, manual, scoring, practice, questions
1:45-3:00 p.m.	Dr. Cort: Sampling procedures
3:00-3:30 p.m.	Coffee break
3:30-6:30 p.m.	Mr. Karl Banks: Vineland Social Maturity Scale - introduction, manual, demonstration, scoring, discussion
6:30-7:30 p.m.	Dinner
7:30-9:30 p.m.	Dr. Lois-ellin Datta: Testing, with special emphasis on tester-child relationships
Night	Reading assignment: review Vineland manual and PSI manual; administer and score PPVT, PSI, BI, and Vineland at least once on own
Saturday, May 14	
8:00-8:45 a.m.	Breakfast
8:45-10:30 a.m.	More demonstration-discussion of tests

EXHIBIT C-2 (Continued)

10:30-11:00 a.m.	Draw-A-Person - introduction, practice, scoring, discussion
11:00-11:45 a.m.	Complete "dry run"; pair testers so that one is "child," the other is tester; find place for testing, establish rapport, test, score, recheck, etc.; "child" also takes role as informant for Vineland and BI (this is all simulation, of course); prepare testers for arrival of children
11:45-12:45 p.m.	Lunch
12:45-1:00 p.m.	Free time
1:00-3:00 p.m.	Pair child with trainee; administration of PSI, PPVT, and DAP
3:00-3:30 p.m.	Debriefing
3:30-4:00 p.m.	Break
4:00-6:15 p.m.	Dr. Charles Dailey: A case study emphasizing adult-adult relationships and problems. Short introductory lecture; discussion of case study (read previously) about reaction by the "establishment" to an outsider; buzz groups; hour discussion--role playing--ratings (small groups); integration and report--entire group
6:15-6:30 p.m.	Free time
6:30-7:30 p.m.	Dinner
7:30 p.m.	More on Vineland, including practice; more on sampling procedures; more on special problems; general review; field assignments and travel arrangements
Sunday, May 15	
9:00-10:00 a.m.	Breakfast
10:00-11:00 a.m.	Clarification of immediate schedules
11:00 a.m.	Leave for airports as appropriate
	General dismissal

EXHIBIT C-3 SCHEDULE FOR SECOND HEAD START TESTER TRAINING PROGRAM, JUNE 8-11, 1966

Wednesday, June 8	
5:30-6:30 p.m.	Mrs. Naomi Henderson: Introductory meeting at PRC, 12th floor conference room; hand out test packets and reading assignments
Thursday, June 9	
9:30-10:30 a.m.	Meeting at PRC, general orientation and background: Dr. Allen Ferguson Dr. H. Russell Cort
10:30-11:00 a.m.	Mr. Charles McDaris: Administration of project
11:00-12:00 noon	Dr. O'Keefe and Mrs. Henderson: Review of nontechnical details in PRC's Head Start testers' manual
12:00-1:00 p.m.	Lunch
1:00-1:40 p.m.	Dr. O'Keefe: Introduction to technical nature of work; overview of all tests
1:40-2:30 p.m.	Dr. O'Keefe: PPVT - manual, scoring, demonstration practice, questions
2:30-3:30 p.m.	Dr. Cort and Dr. William Commins: Sampling procedures
3:30-4:00 p.m.	Break
4:00-6:00 p.m.	Dr. Lois-ellin Datta: Testing, with special emphasis on tester-child relationships and factors
6:00 p.m.	Adjourn, with study assignment: PPVT and manual; PSI and manual; Vineland manual--focus on eight categories; Head Start testers' manual; administer and score two PPVT's; prepare for test on Vineland's eight categories

EXHIBIT C-3 (Continued)

Friday, June 10

9:00-9:30 a.m.	Check-in at Center of Adult Education, University of Maryland, Adelphi Road and University Boulevard, College Park, Maryland (phone 301-779-5100)
9:30-11:00 a.m.	Dr. O'Keefe: PSI - manual, practice, scoring, questions DAP - administration
11:30-12:00 noon	Dr. O'Keefe: Complete review of test procedures and forms, and preparation for children's arrival after lunch
12:00-1:00 p.m.	Lunch
1:00-3:30 p.m.	Test children
3:30-3:50 p.m.	Break
3:50-4:10 p.m.	Dr. O'Keefe: Introduction to Vineland Social Maturity Scale
4:10-6:30 p.m.	Mr. Karl Banks: Demonstration of Vineland and scoring; discussion
6:30-7:30 p.m.	Dinner
7:30 p.m.	Reconvene to receive study assignment: read case study; give and score Vineland twice; give PSI and PPVT once; mark and correct two BI's; simulate a DAP; prepare one or more complete simulation sets of data, organized as per testers' manual

Saturday, June 11

8:00-9:00 a.m.	Breakfast
9:00-9:30 a.m.	Submit data prepared Friday night
9:30-12:00 noon	Dr. Charles Shaffer: A case study emphasizing adult-adult relationships
12:00-1:00 p.m.	Lunch

EXHIBIT C-3 (Continued)

1:00-3:30 p.m.	Test children from a local Head Start class: Administration of PSI, PPVT, DAP, BI (simulate), Vineland (simulate), and record keeping
3:30-4:30 p.m.	Score data; break
4:30-5:30 p.m.	Debrief testing experience
5:30-6:30 p.m.	Mrs. Henderson: Discuss record keeping Dr. O'Keefe: Review of all test data submitted
6:30-7:30 p.m.	Dinner
7:30 p.m.	Travel assignments and completion of all plans Adjourn

EXHIBIT C-4 EXTRACT FROM PRC'S TESTER'S HANDBOOK

Administering Tests

1. General Procedures

The Preschool Inventory will help you establish rapport with the child and, in any case, must be given before the PPVT. You may find it quite natural to start giving the PSI when you first meet the child by asking him his name and age. The following suggestions may be helpful.

a. To establish Rapport With a Child:

(1) Say: "We are going to play a game."

(2) Give him some of the props (cars, checkers).

(3) Ask him questions about the props (i.e., get him talking).

(4) Use random, intermittent reinforcement for incorrect as well as correct responses. Statements such as "that was a good answer," are good. If he makes no answer even after gentle prodding, move on to the next one, saying: "Let's try another one."

b. To Interest a Bored or Distracted Child

(1) Make him believe he is doing a very important thing by helping you.

(2) Keep your voice interesting; do not talk in a monotone.

(3) Tell him that you will be finished soon.

(4) If he seems seriously uninterested, perhaps a break from testing will be necessary. Employ the principle of playing simple games or singing etc., to employ a "change of pace." Return to the testing after 5 minutes.

c. To Make a Child Stop Crying

(1) Gain his confidence by offering him sympathy and diverting his attention from himself to you or inanimate objects.

EXHIBIT C-4 (Continued)

(2) Return child to class and test him later (if possible) if he seems seriously disturbed about the testing situation.

(3) Reexamine your appraoch with the children if many are reacting this way.

THE TESTS MUST BE GIVEN IN THE FOLLOWING ORDER:

2. Preschool Inventory (PSI) - Used directly with child

a. At first you will find it necessary to give the test verbatim from the manual. As you become familiar with the acceptable answers you may use the test form alone. In any case, be sure to have the manual with you. It is important to phrase the questions exactly as stated in the manual. Do not reword a question if a child has trouble understanding the thought; repeat it once.

b. The difference between a "2" and a "1" response (items 43-47) depends on whether the child describes a general function of this person in society rather than a specific duty or job.

c. Refrain from accenting the key word in a sentence; it offers verbal clues. Practice privately using the actual questions to train yourself away from this tendency.

d. You may notice a tendency to become more lenient after testing several children, particularly if you notice that many children make the same kind of responses. This is very easy to do--try to avoid this.

e. Speak slowly and clearly but do not overaccent words. Keep your normal speaking voice throughout the testing. Avoid becoming "sing-song."

3. Peabody Picture Vocabulary Test (PPVT) - Used directly with the child

a. Read the directions on page 7 of the manual to the child verbatim.

EXHIBIT C-4 (Continued)

b. Avoid preceding a stimulus word by an article (a, an, the) or converting words to plurals as this may provide a cue to the child.

c. Always secure a response from the child. Do not record an answer of "no response." You may encourage him to guess.

d. In order to encourage a response, start the test with "point to," "what number is," or "where is," preceding the stimulus word. As the test proceeds you may drop the introductory phrase.

e. After giving test, check test book for smudges, and erase all marks.

f. Be sure the child's 11-digit identification number is on the upper righthand corner of each piece of his data.

4. Draw-A-Person (DAP) - Used directly with child

a. DAP is a measure of mental maturity which is based on the degree of complexity of the child's pencil drawing of a person.

b. Make child understand that you want him to draw the best possible person. In your instructions simply say: "I want you to draw a person." Then in answer to a question of what kind of person, reply: "The best one you know how," or "Any person you want to draw."

c. Do not give clues as to details, such as "What about arms and legs?"

d. Remember to give the child a pencil and not a crayon.

e. Encourage the child to draw something.

5. Vineland Social Maturity Scale - Information collected from the head teacher (or other staff member who knows the child well) via interview

a. Learn as much as possible about the child before proceeding with the test, such as age, schooling, general ability, and home environment. This will help you evaluate the information given to you.

b. Begin questioning well below the anticipated final score.

EXHIBIT C-4 (Continued)

c. Quiz the teacher in a sympathetic manner; ask does Johnny usually do so and so rather than can he do so.

d. After obtaining as much information as possible from the teacher, make YOUR OWN scoring judgment.

e. You should present the various items and groups of items by category, as explained in the manual.

f. FAMILIARIZE YOURSELF WITH THE EIGHT GENERAL CATEGORIES AND THEIR ABBREVIATIONS.

g. The most serious difficulty in giving this test is failing to get sufficient detailed information to make a judgment. Do not become impressed with random statements such as "Johnny is such a sweet child," "Sally is such a good little girl," or "Jimmy is a hellion." These statements are meaningless.

6. Behavior Inventory (BI) - Used by the teacher

The head teacher should fill out the Behavior Inventory for each child tested in her class. She should give her general impression of the child, as there are no right or wrong ratings.

EXHIBIT C-4 (Continued)

Scoring Test Forms

1. Preschool Inventory (PSI)

a. All items except Nos. 43-47 are scored as either correct (1 point) or incorrect (0 points). No distinction is made between a wrong answer and no answer.

b. Questions 43-47 are scored either 2 or 1, depending on the level of abstraction as explained in the PSI manual (page 9).

c. Add the scores for each subtest and the total test and write them on the front of the form.

d. Determine the child's age in months.

2. Peabody Picture Vocabulary Test (PPVT)

a. The basal score is determined by working forward on the test until the child makes his first error. If he has not made eight consecutive correct responses prior to this first error, drop back to the starting point and work backward until he has made a total of eight consecutive correct responses.

b. The ceiling is reached when six errors have been made in the last eight items presented; count the last item presented as the ceiling.

c. After subtracting the errors from the ceiling, the mental age and I.Q. can be determined by consulting the appropriate table in the manual. When the I.Q. is below about 55, see page 49 in the manual.

3. Draw-A-Person (DAP) Test

a. Do not score DAP.

b. Small children often make the same kind of drawing. You will become familiar with some of the similar methods of representing body characteristics. For example: small children often represent the two arms, but they attach them to the head. If you watch carefully while they are doing this, you will realize that they are arms and not ears.

EXHIBIT C-4 (Continued)

BE SURE TO RECORD INFORMATION CONCERNING ADMINISTRATION TIMES OF TESTS ON PSI, PPVT, AND VINELAND.

4. Vineland Social Maturity Scale

a. One of five scores may be given for each item on the Vineland Social Maturity Scale:

(1) (+) for items habitually performed (full credit)

(2) (+NA) for items which the child has had no opportunity to perform (full credit if it falls within the range of otherwise continuous scores, no credit if it falls within range of otherwise continuous minus scores, and half credit if it falls within the intermediate range)

(3) (±) for items in the transitional or emergent state (half credit)

(4) (-) for items never or rarely successfully performed (no credit)

(5) (-NO) for items which the child has had no opportunity to perform and probably would not be able to perform successfully

b. The highest continuous plus score for all items is considered the basal score; the allowance made for lack of opportunity is noted above. Plus credit is assumed for all items below the basal.

c. A completed form should show at least two consecutive minus scores in each of the eight categories.

d. The total score is obtained by adding to the basal score the additional credits scattered beyond the basal score.

e. The total score is then converted to an age equivalent (social age, or SA) by checking the table on page 28 in the manual.

f. Determine the Social Quotient by using this formula:

$$\text{Social Quotient (SQ)} = \frac{\text{Social Age (SA)}}{\text{Life Age (LA)}} \times 100$$

Note: Life age means chronological age.

EXHIBIT C-4 (Continued)

g. Finally, complete the face sheet of the record form by recording the information obtained (basal score, additional points, total score, social age--which is called "age equivalent" on the record form--and social quotient).

5. Behavior Inventory (BI)

a. Be sure all identification numbers are filled in by the teacher for each child's inventory.

b. If necessary, read instructions on inventory with teacher before she completes forms.

c. Score the BI by using the cardboard scoring key (BI Corrector).

APPENDIX D

TESTS

This appendix provides additional detail on tests to supplement information presented in Sections III, IV, and V. Topics covered are:

- Subjective critiques of the primary test instruments, based mainly on testers' comments.
- Spanish translations of tests.
- Data on test administration time for PPVT and PSI.
- Copies (English and Spanish) of the Family Information Form (CAP-HS Form 46) and Behavior Inventory.

A. Discussion of Tests

There is currently considerable interest among educators in determining and/or developing measures of program effectiveness and child development--especially measures appropriate for use with so-called culturally disadvantaged children. This section presents an informal discussion of tests and measures used in this study, based primarily on the testers' experiences and opinions. While no attempt was made to elicit systematic observations from testers about the tests, most testers did in fact make at least some written comments on specific test items or frequently encountered test problems. This section is not intended as a criticism of the tests selected for use, their standardization procedures, or their overall applicability to young culturally deprived children. Its purpose is simply to point out test items or requirements which frequently posed problems in testing.

1. Peabody Picture Vocabulary Test (PPVT)

The testers did not encounter any significant difficulties in administering the PPVT.[1] Once the simple pointing response was established, it was relatively easy to proceed through the testing and maintain rapport.

---

[1] The tester shows the child a page with four pictures and says, "Show me...(ball, etc.)."

There were, however, a few items on the test that elicited unusual answers from the children frequently enough to be noted here.

Item 15	The test stimulus word is "pulling." The correct picture is that of a child pulling a wagon, but children often selected the picture of a girl about to knock on a door or open it by the knob.
Item 24	The stimulus word is "baking." The correct picture is that of a girl placing a cake in an oven, but children frequently pointed to the picture of a baby eating.
Item 30	The test word is "time" and the correct picture is a clock. However, many children spontaneously began talking about the picture of two cars crashing.
Item 38	The test word is "barber," but many children pointed to the picture of a queen on a throne. Some testers commented that especially non-white children in urban areas have been responding to a hairdresser who sets a woman's hair, rather than a man who cuts hair.
Item 44	The stimulus word is "cash," but many children seemed to hear this as "crash" and pointed to a picture of a broken dish.

Some testers commented that children often seemed to enjoy the PPVT. Since it was always administered after the PSI, which some children found discouraging, in testers' opinions, there seemed to be a contrast in ease that appealed to the children. Some testers in fact suggested that in the future the PPVT should be given first as an ice-breaker. Exactly how the effects of this observation relate to the test-anxiety hypothesis discussed in Section V is difficult to say.

2. The Pre-School Inventory (PSI)

The Pre-School Inventory, developed by Dr. Bettye Caldwell and Mr. Donald Soule of the State University of New York, consists of four subtests designed to assess personal-social responsiveness, associative vocabulary, and numerical and sensory concept activation.[1] In addition to testers' observations, the item analysis in Exhibit 35, Section IV, will serve as a basis for this discussion; Exhibit 35 lists, item by item, for both age and program duration, the percentage of children who responded correctly.

Concerning the general procedures for administering the PSI, many testers commented that items 1 and 2, which asked the child for his first and last name, were troublesome because, to avoid asking the question twice, the tester had to be careful not to ask "What is your name?" while establishing rapport upon first meeting the child (even though it is a natural question with which to begin a conversation). Some testers noted that some children seemed baffled by such an obvious question. Many testers commented that item 3, which asks the child's age, would make a good, natural first question to ask.

In addition, testers frequently noted problems stemming from the use of various test materials. The brightly colored plastic cars, for example, were often a novelty for the children, who seemed to view them as attractive toys. The child's main interest in many cases was playing, and problems often resulted when testers attempted to position the cars for testing purposes. Exhibit 35 shows that PSI items 19 through 26, which used cars as part of the test procedure, proved especially difficult for the children.

While checkers caused fewer problems, their use as test props also presented certain difficulties. Many children appeared to focus their attention on the design of an eagle or crown imprinted on the checker and could not "imagine" the checker as a train car. As Exhibit 35 indicates, on items 27 and 28, in which a line of checkers became a train, very few children of any age responded correctly.

Although testers had no difficulty in following correct procedures for scoring the PSI, their comments on items 43 through 47[2] suggest that the rating or score assigned to a child's response may be unjustly low. The items themselves call for a verbal response; the child receives two points if he makes a generalization and one point if he describes a specific function (or functions) of the person or concept in question. Exhibit 35 shows that only a very few children (usually less than 10 percent) scored two points on any item. Testers' reports, however, indicated that many children did make a "generalization type" of response, but that it deviated greatly from acceptable answers listed in the manual. For example, in item 46, which asks what a father does, few children answered with the sample responses listed in the manual: "Takes care of family--earns money for family." A large number of children, even with a prompting, answered: "He drinks beer; he sleeps; he watches TV; he goes out." Such responses suggest that the child recognizes the concept in question.

---

[1] Caldwell, Bettye and Donald Soule. The Pre-School Inventory, 1966.

[2] Items 43 through 47 ask what the following people do: dentist, policeman, teacher, father, and mother.

The test directions used for items 29 through 33[1] presented considerable difficulty. A correct response required both a verbal statement and a gestural (hand-motion) response. Testers felt that more than one example should have been included in the test directions. Further, some testers felt that the verb "go," (used in the sense of "What is the direction of...?") may have been misleading to many children. For example, a tester noted that a child answering item 32 ("Which way does a phonograph record go?") responded by humming a tune. When asked "How else does it go?" he responded by humming a different tune--a creative response, but an "incorrect" one.

Certain individual PSI items presented specific difficulties for the children, according to testers' notes. Often, a look at Exhibit 35 indicates a high correlation between individual testers' comments on specific items and overall results by children on that item. For example, for item 11, the tester points to his or her own knee asking "What's this?" Testers commented that often the child thought that a female tester was pointing to her hosiery, and would reply "stocking" or "nylon." In the case of male testers, children often replied "trouser" or "trouser leg." Exhibit 35 shows that item 11 does discriminate among age levels, but the overall percentage of correct responses is still very low.

---

[1] Items 29 through 33 ask which way the following go: saw, elevator, ferris wheel, phonograph record, and waterfall.

For item 17, which asks the child to stand up and face the door, problems were posed when a school or test room did not have a door. For example, in a school in a large urban area, there was a folding accordion wall rather than an actual door. In that school, none of the 15 children tested could respond correctly. Perhaps a more body-oriented question such as "Stand up and put your hands on your ears" would fulfill the test purpose better.

Items 19 through 26[1] required the use of brightly colored plastic cars and paper boxes. Items 22 and 25, as seen in Exhibit 35, proved especially difficult at all age levels. It seemed at times, according to the testers' comments, that the children were "snowed under" by all that was desired of them. Further, many testers mentioned that, because the tester was required to return all props to their original positions following each of those items, the children seemed to have the feeling that the testers were "undoing" what the children had done.

It can be seen in Exhibits 33 and 34 that Subtest 2, which was designed to measure associative vocabulary, contains fewer correct answers than any other subtest. Items 27 and 28 ("What do we call the first car, the one that pulls the train?" and "What do we call the last car on the train?") used checkers to represent a train, and clearly presented a problem for children at all age levels. The possible interference of the checkers has already been mentioned. A second hypothesis is that children may have been unable to abstract the concept of train when viewing the concrete representation of the checkers. In any case, the level of difficulty noted in Exhibits 33 and 34 seems significant enough to merit further investigation.

Items 35 through 36 ask questions about the seasons of the year.[2] Again, as Exhibit 35 points out, few children at any age level answered correctly. Testers noted that children often gave an answer such as "5 o'clock." Perhaps vocabulary might be studied to determine whether the choice of terms used interfered with the children's comprehension of the test question.

---

[1] These items ask the child to place various colored cars on, under, in, and behind different sizes and colors of boxes.

[2] Items 35 through 37 ask what time of year is hottest and coldest, and what time of year it is now. See discussion of this subject in Section V.

Subtest 3, designed to measure numerical concept activiation, discriminates rather well among the different age levels. However, item 51, which asks how many toes the child has, was very difficult, even for 6-year-olds. Item 56, which asks how many wheels a rowboat has, also received very few correct answers.

For Subtest 4, designed to measure sensory concept activation, testers commented that for items 81 and 82 (which ask which of eight crayons is the color of the sky and which is the color of the night) many children explained, correctly, that the sky is white or gray (clouds, smog, etc.) but these colors were not among the choices available.

The PSI covers a variety of concepts in a simple and direct manner. Like any instrument in its early stages of development, it is not without problems. The item analysis in Exhibit 35 and the informal comments made by testers may offer suggestions for further improvement of the measure.

3. Vineland Social Maturity Scale (VSMS)

The Vineland Social Maturity Scale (VSMS) is an interview schedule which follows a detailed outline designed to evaluate the child's maturity in eight areas of development, including social and personal.

Of all the measures, this was the most difficult to administer because the scale is based on the assumption that the respondent will be a parent or someone who has actually lived with the child. However, testers interviewed teachers or teacher aides, many of whom admitted their inability to provide reliable information for certain questions.[1] In such cases, testers requested the respondent to make a judgment as to whether or not the child would be "capable" of the behavior in question. One tester commented that too often some responses from some respondents seemed to be a combination of conjecture, opinion, and fiction! Sometimes it seemed to testers that a teacher tended to describe all children in her class in very similar terms.

---

[1] The following items frequently presented difficulty for respondents:
No. 48 (helps at little household tasks)
No. 64 (bathes self assisted)
No. 65 (goes to bed unassisted)
No. 72 (does routine household tasks)
No. 74 (bathes self unaided)

Testers often noted that the initial rapport established with the teacher or aide greatly influenced the amount of information obtained during the VSMS. Testers reported that some respondents were uncooperative and that their unwillingness to spend time for interviews definitely limited their responses made on the VSMS. Sometimes, when a tester found it necessary to administer several Vinelands to one teacher, the respondent developed a tendency either to answer all questions in a general manner or to rate different children rather uniformly on all items.

4. Behavior Inventory (BI)

Testers' comments revealed the existence of several difficulties with the Behavior Inventory (BI). Incomplete Behavior Inventories were often returned to the tester. Teachers repeatedly left items 1, 27, 41, and 44 unchecked--a fact which strongly suggests that these items were overlooked due to poor format design and uneven layout. Since any incomplete BI was useless as a research measuring instrument, it was necessary to mail a large number of these forms back to the teachers for completion.

Tester observations frequently quoted teachers and aides as commenting that most of the children were alike. Consequently, it was often noted that staff members tended to fill out BI's rather rapidly. Thus, despite the fact that testers were careful to give the BI to a staff member who knew and worked closely with the child, the accuracy of the responses may in some cases be questioned.

Testers also noted that some teachers, and many of the aides, seemed confused by certain vocabulary items. An examination of the form reveals a rather sophisticated level of vocabulary, such as "imperturbable" in item 23, "lethargic" in item 46, and "usurps" in item 16. Tester data do not attempt to include any record of the respondent's understanding of particular items, but in view of the wide range of educational levels reported for staff members who completed the BI, it would be interesting to see if the vocabulary or general language style could significantly affect responses made.

Further difficulties arose because of the complexity or ambiguity of test items. In many cases, teachers complained that only a part of the item was applicable to a child. For example, item 49 ("Approaches new tasks timidly and without assurance; shrinks from trying new things") mentions two characteristics; a child may be very timid, but nevertheless, he will still try something new. Item 32 ("Is reluctant to use imagination; tends not to enjoy 'make believe games',") presents the same rating problem, for while the object is to measure a negative behavioral trait, many children who have little interest in make-believe games still possess and use their imaginations. Item 7 ("Often keeps aloof from others because he is uninterested, suspicious, or bashful") asks for a single rating on three separate attributes which are not synonymous.

In other items, the ambiguity of terms makes it possible for respondents to interpret key words differently. For example, in item 47 ("Has a tendency to discontinue activities after exerting a minimum of effort") it is difficult to know by what criteria teachers judged a child to have exerted a "minimum of effort." Item 12 ("Is rarely able to influence other children by his activities or interests") also has a sense of ambiguity. Here again, how does one determine or measure a child's influence on others?

This absence of clearly defined criteria may also have affected the child's score where an item, which is rated as positive on the Behavior Inventory, was interpretated by the respondent as a negative characteristic. One example is item 8 ("Defends or praises his own efforts"). In some cases there is only a fine line of difference between the positive and negative characteristic being measured. Compare, for example, items 5 and 19.[1]

The significance of these complexities and ambiguities has not been determined. However, while such factors may be insignificant in terms of overall validity, their appearance suggests a need for considerable further refinement of the inventory.

---

[1] These items read as follows: "Talks eagerly to adults about his own experiences and what he thinks" and "Is excessive in seeking the attention of adults."

B. Spanish Test Translations: PSI, PPVT, and DAP

Spanish translations of the Pre-School Inventory (PSI), the Peabody Picture Vocabulary Test (PPVT), and the Draw-A-Person (DAP) test were prepared by two testers who were fluent in the Spanish language. Although the translations were primarily based on Spanish spoken in Southern Texas, (i.e., "Mexican Spanish" as opposed to "Castilian Spanish") they were found to be fairly consistent with Spanish used in other areas such as the West Coast. If, however, a particular word was not commonly used in a given locale, the tester substituted a more familiar Spanish word, and noted the substitution on the child's test record. The testers began administering the first test (PSI) in English and transferred to the Spanish language as soon as it appeared that the child was having difficulty in understanding the tester's English language.

The general and detailed instructions for administering the PSI, PPVT, and DAP in Spanish follow.

1. General Instructions to Testers for Administering the PSI, PPVT, and DAP in Spanish

GENERAL NOTE: You should consult bilingual people at your test center in order to pick up any regional differences in the Spanish. These translations were based on Spanish spoken in South Texas, and may vary from that spoken in New York, California, or Miami. If you find that a word used on the test is not one used in your region, use the latter, and indicate the change on the test.

PSI - In the brief period before the testing begins, try to discern in which language the child is most comfortable. If the children are shy, this may be difficult to do. Also, since you will be looked upon as a teacher, and since English is usually spoken in the classroom, the child may feel he has to speak in English. Begin the test in English, and carry the child as far as you can in that language. If he responds incorrectly, or appears reluctant or frightened to respond, ask the question a second time in Spanish and mark him according to his second response. After the first direction in Spanish, you should make every subsequent direction first in English, followed by the Spanish. Do not pause to allow for a response between the two. Indicate this point by placing an S in the margin of the answer sheet. Responses: If, while working exclusively in English, the child responds in Spanish, indicate this response by placing an S inside the box on the answer sheet. Once you have begun using English/Spanish, indicate the language in which the verbal response was given by placing either an E or an S inside the box on the answer sheet.

PPVT - Give Examples A, B, and C in English. If all three are recognized, begin with Item 1 and continue as far as you can in English. The first time a child misses a word, repeat the word in Spanish and indicate this by placing an S in the margin. (If the child recognizes it in Spanish, give him credit for it.) Now continue the test giving the English word first, followed by the Spanish word. Do not pause to allow time for a response between the two words. However, if you notice the child moving his hand in response to the English word, indicate this by placing an E in the margin.

DAP - Direction should be: "Haz un retrato de una persona, el mejor que puedes."

2. Specific Instructions

a. PSI

1. Cuál es tu nombre?
2. Cuál es tu apellido?
3. Cuantos años tienes?
4. Cuándo es tu cumpleaños?
5. Muéstrame tu ojo.
6. Muéstrame tu' cuello.
7. Muéstrame tu hombro.
8 Muéstrame tu talón.

9. - 12. (Qué es esto? - Cómo se llama? - Qué más?)

9. Oreja
10. Dedo
11. Rodilla
12. Codo

Muy bién. Quiero que hagas unas cosas para mi.

13. Levanta la mano.
14. Muévete.
15. Di Hola en voz alta.
16. Di Hola en voz suave.
17. Ahora, levantate y haga frente a la puerta.
18. Ahora, brinque.

19. - 26. - Muy bién. Ahora, sientate en la silla.

19. Pon el carro rojo encima de la caja negra.
20. Pon el carro azul debajo de la caja verde.
21. Pon el carro amarillo encima de la caja chiquita.
22. Pon un carro en la caja mediana.
23. Pon todos los carros en un lado de la mesa y todas las cajas en el otro lado de la mesa.
24. Pon tres carros en la caja grande.
25. Pon dos carros detras de la caja en el medio.
26. Dame todo a mí.

27. - 28. Imagínate que este es un tren. Sabes lo que es un tren? Tu sabes, tiene muchos carros, uno trás de otro, así.

27. Sabes lo que llamamos al primer carro, el que estira el tren?
28. Como se llama el ultimo carro de un tren?

29. - 33. - Sabes lo que es un columpio? Tu sabes como se mueve un columpio -- arriba y abajo y por el atras y por el frente.

29. Bueno, como se mueve un serrucho?
30. Como se mueve un ascensor?
31. Como se mueve una montaña de carnaval?
32. Como se mueve un disco?
33. Como se mueve una cascada?
34. Cuando desayunamos?
35. En cuál tiempo del año hace más calor?
36. En cuál tiempo del año hace más frio?
37. En cuál tiempo del año estamos ahora?
38. Si quieres encontrar un león, dónde lo buscarías?
39. Si quieres comprar gasolina, a donde irías?

40. Si estarias enfermo(a) a quien verías?
41. Si quieres encontrar barco, dónde lo buscarías?
42. Si quieres leer algo, que harías?
43. Que hace un dentista?
44. Que hace un policia?
45. Que hace un maestro?
46. Que hace un papá?
47. Que hace una mamá?
48. - 56. - Cuántos hay? De todo, cuántos hay?
48. Cuántos ojos tienes?
49. Cuántos narizes tienes?
50. Cuántos manos tienes?
51. Cuántos dedos en los pies tienes?
52. Cuántas ruedas tiene un carro?
53. Cuántas ruedas tiene una bicicleta?
54. Cuántas ruedas tiene un bicicleta chiquita?
55. Cuántas ruedas tiene una carretilla?
56. Cuántas ruedas tiene un barco de remo?
57. Cuenta en voz alta.
58. Cuantas esquinas tiene este papel?
59. Cuál tiene más damas?
60. Cuál tiene más damas?
61. Cuál tiene menos damas?
62. - 66. Pon estas damas en una linea.
62. Dame el del medio.
63. Dame el primero.
64. Dame el ultimo.
65. Dame el segundo.
66. Dame el penúltimo.
67. - 70. - Ahora, quiero que harás unos dibujos. Haz uno así.
71. - 73. - Cuál se parece más un(a)____________________.
Cual de los dibujos es eso?
71. Rueda.
72. Carpa.

73. Palo.
74. Cuál es más mayor, una pelota o una bicicleta?
75. Cuál es más mayor, un arbol o una flor?
76. Cuál es más despacio, un carro o una bicicleta?
77. Cuál pesa más, un ladrillo o un zapato?
78. Cuál pesa más, una pluma o un tenedor?
79. De que color es?
80. De que color es?
81. Cuál de estos es el color del cielo?
82. Cuál es el color de la noche?
83. Pinta el circulo amarillo.
84. Pinta el cuadro morado.
85. Pinta el triangulo naranja.

b. PPVT

1. table - mesa
2. bus - bus
3. horse - caballo
4. dog - perro
5. shoe - zapato
6. finger - dedo
7. boat - barco
8. children - niños
9. bell - campana
10. turtle - tortuga
11. climbing - subiendo
12. lamp - lámpara
13. sitting - sentado
14. jacket - saco
15. pulling - estirando
16. ring - anillo
17. nail - clavo
18. hitting - pegando
19. tire - llanta

20. ladder - escalera
21. snake - víbora
22. river - río
23. ringing - sonando
24. baking - cocinando
25. cone - cono
26. engineer - maquinista
27. peeking - asomando
28. kite - güila
29. rat - ratón
30. time - hora
31. sail - vela
32. ambulance - ambulancia
33. trunk - baúl
34. skiing - esquiando
35. hook - anzuelo
36. tweezers - pinzas
37. wasp - avispa
38. barber - barbero
39. parachute - paracaidas
40. saddle - silla de caballo
41. temperature - temperatura
42. captain - capitán
43. whale - ballena
44. cash - dinero
45. balancing - balanciando
46. cobweb - telaraña
47. pledging - jurando
48. argument - pleito
49. hydrant - mono de agua
50. binocular - largavistas
51. locomotive - locomotora
52. hive - colmena

53. reel - carrete
54. insect - insecto
55. gnawing - royendo
56. weapon - arma
57. bannister - baranda
58. idol - ídolo
59. globe - globo
60. walrus - morsa
61. filing - archivando
62. shears - tijeras
63. horror - horrór
64. chef - cocinero
65. harvesting - cosechando
66. construction - construcción
67. observatory - observatorio
68. assistance - asistencia
69. erecting - haciendo
70. thoroughbred - de sangre pura
71. casserole - casuela
72. ornament - adorno
73. cobbler - zapatero
74. autumn - otoño
75. dissatisfaction - descontento
76. scholar - escolar
77. oasis - oasis
78. soldering - soldando
79. astonishment - sorpresa
80. tread –
81. thatched - techado con paja
82. jurisprudence - jurisprudencia
83. sapling - renuevo
84. arch - arco
85. dwelling - residencia
86. lubricating - aceitando

87. pedestrian - peatón
88. vale - cañada
89. jubilant - jubiloso
90. laden - cargado
91. pursuit - persiguiendo
92. goblet - cáliz
93. rodent - roedor
94. confiding - contando
95. reclining - acostado
96. frisking - esculcando
97. moat - mote
98. salutation - salutación
99. barrier - barrera
100. foal - potro
101. incandescent - incandescente
102. cornucopia - cornucopia
103. ascending - ascendiendo
104. summit - ápice
105. caster - rodadillo
106. lobe - lóbulo
107. patriarch - patriarca
108. sampler - bordado
109. ingenious - ingenioso
110. repose - reposo
111. constrain - detener
112. tangent - tangente
113. sconce - candelabra de pared
114. hoary - canoso
115. pendant - pendiente
116. prodigy - prodigio
117. casement -
118. quiescent - tranquilo
119. talon - garra
120. chevron - galón

c. DAP

Directions should be: "Haz un retrato de una persona, el mejor que puedes."

C. Time Required to Administer PPVT and PSI

Exhibit D-1 provides information in bar graph form on the mean number of minutes spent by testers in administering the PPVT and the PSI to 3-, 4-, 5-, and 6-year-old children. The exhibit shows that, as might be expected, the older the child, the more time was required to administer the PPVT. (The PPVT is open-ended, and older children usually cover more test items.) For the PSI, however, 4- and 6-year-olds on the average required about the same amount of time (23.8 and 23.5 minutes), 5-year-olds required the least amount (22.6 minutes), and 3-year-olds needed the most time (25.1 minutes). (On the PSI, the number of items remains constant, and older children tend to go through them a bit more quickly than children as young as 3.)

The mean administration time for the PPVT is given as 10 to 15 minutes in the PPVT test manual.[1] Although PRC found that the mean times ranged from 8.5 to 11.0 minutes (with standard deviations ranging from 3.3 to 4.4 minutes), the PRC study dealt only with relatively young children, while the PPVT manual deals with children up to age 18.

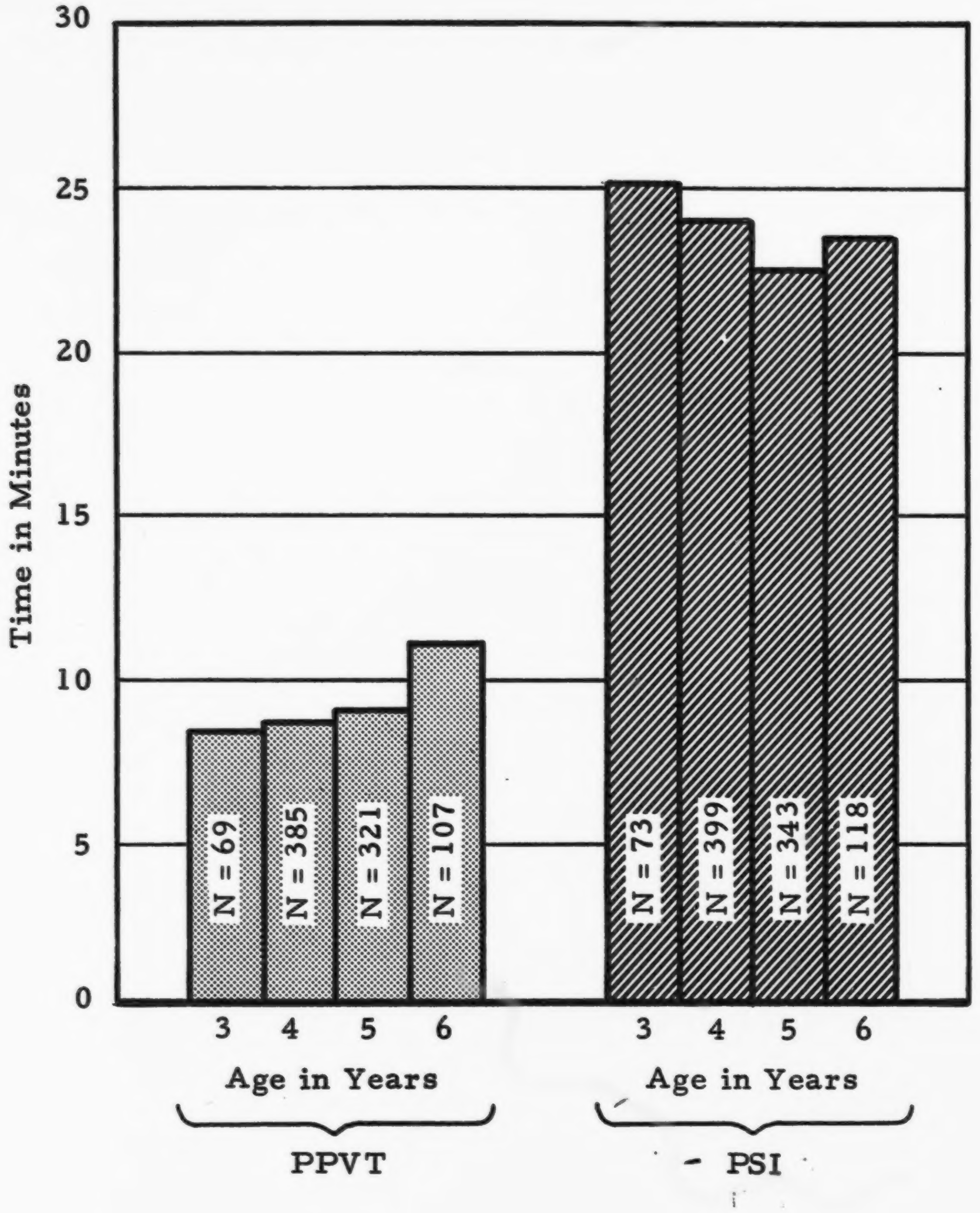
According to the PSI manual, the inventory "takes no more than 15 minutes to administer to most children."[2] This figure differs somewhat from the required time as reported by PRC testers, since mean times (for different age groups) ranged from 22.6 to 25.1 minutes (with standard deviations ranging from 6.6 to 8.5 minutes).

In terms of planning time allowances for future testing of Head Start children, the time differences between the various ages from 3- to 6-year-olds would probably be insignificant. Realistically, about 10 minutes should be allowed per child for the PPVT, and about 25 minutes per child for the PSI.

---

[1] Dunn, Lloyd, Expanded Manual, Peabody Picture Vocabulary Test. Minneapolis, Minnesota: American Guidance Service, Inc., 1965, p. 5.

[2] Caldwell, Bettye and Donald Soule. The Pre-School Inventory (Manual), 1966, p. 3.



Note: (1) N's (sample sizes) for each age differ slightly between PPVT and PSI because test administration time was sometimes not recorded by testers.

EXHIBIT D-1 MEAN TIME IN MINUTES REQUIRED TO ADMINISTER THE PPVT AND PSI TO THREE-, FOUR-, FIVE-, AND SIX-YEAR-OLD CHILDREN(1)

D. Copies of Family Information Form and Behavior Inventory

The English and Spanish versions of the Family Information Form (CAP-HS Form 46) and the Behavior Inventory are presented in Exhibits D-2, D-3, and D-4, respectively.

EXHIBIT D-2 ENGLISH VERSION OF FAMILY INFORMATION FORM (CAP-HS FORM 46)

PRC R-886
D-21

OFFICE OF ECONOMIC
OPPORTUNITY

EXECUTIVE OFFICE OF THE PRESIDENT
WASHINGTON, D.C. 20506

BUDGET BUREAU NO. 116-R052
APPROVAL EXPIRES JUNE 30, 1967

Dear Parent or Guardian:

The Head Start School in which your child is enrolled is partially financed by the Office of Economic Opportunity in Washington, D.C. In order to explain the program to the Congress and the general public, the Office of Economic Opportunity needs some facts about Head Start families.

Your answers to the questions on this form will be combined with thousands of others from all over the United States. This should give us information about the families which are enrolling children in the Head Start program.

If you have trouble answering any questions on the form, please ask for help with it at the Head Start Child Development Center. When you finish, please mail the form right back in the enclosed postage-paid envelope.

Sincerely yours,

Jule M Sugarman

Jule M. Sugarman
Associate Director
Project Head Start

If the return envelope is
missing, mail this form to:

Project Head Start (Summer)
Census Operations Office
1201 East Tenth Street
Jeffersonville, Indiana 47130

CAP-HS FORM 46 (6-20-66)

PRC R-886
D-22

The following questions are about all the people with whom the child in the Head Start school usually lives. Answer them by marking an "X" in the correct box or entering a number in the space provided.

LIST OF PEOPLE		NUMBER AND AGES OF PEOPLE
1. Is there a father, step-father, or foster father living with the child?	1 ☐ Yes → 2 ☐ No	**1a. How old is he?** 1 ☐ Less than 21 years old 2 ☐ 21-54 years 3 ☐ 55-64 years 4 ☐ 65 or older
2. Is there a mother, step-mother, or foster mother living with the child?	1 ☐ Yes → 2 ☐ No	**2a. How old is she?** 1 ☐ Less than 21 years old 2 ☐ 21-54 years 3 ☐ 55-64 years 4 ☐ 65 or older
IF THERE IS NO FATHER OR MOTHER AT HOME: **3. Who is the child's guardian (the person chiefly responsible for the child's care)?**	1 ☐ A brother, uncle, grandfather, or other male relative → 2 ☐ A sister, aunt, grandmother, or other female relative → 3 ☐ A man – not related → 4 ☐ A woman – not related →	**3a. How old is the guardian?** 1 ☐ Less then 21 years old 2 ☐ 21-54 years 3 ☐ 55-64 years 4 ☐ 65 or older
4. Are there any brothers or sisters living at home? *(Do not count a brother or sister who is acting as the child's guardian.)*	1 ☐ Yes → 2 ☐ No	**4a. How many of them are:** Number Under 6 years old? .... ☐ 6-15 years? .... ☐ 16-21 years? .... ☐ 22 years or older? .... ☐
5. Are there any other relatives (such as grandparents, cousins, and so on) who live in the same house as the child? *(Do not count a relative who is acting as child's guardian.)*	1 ☐ Yes → 2 ☐ No	**5a. How many of them are:** Number Under 6 years old? .... ☐ 6-15 years? .... ☐ 16-21 years? .... ☐ 22-54 years? .... ☐ 55-64 years? .... ☐ 65 years or older? .... ☐
6. Are there any other people who live in the same home but are NOT related to the child (nor acting as the child's guardian)?	1 ☐ Yes → 2 ☐ No	**6a. How many people live in the same house and are NOT related to the child?** Number ______________

If the child is living with parents, the following questions are about them.
If the child is NOT living with parents, the following questions are about his guardian.

7. Please mark an "X" in the box which shows the highest grade of school the parents or guardian completed.

Father or male guardian

1 ☐ No school
2 ☐ 1 to 3 years
3 ☐ 4 to 6 years
4 ☐ 7 to 8 years
5 ☐ 9 to 11 years
6 ☐ High school graduate
7 ☐ Some college or college graduate

Mother or female guardian

1 ☐ No school
2 ☐ 1 to 3 years
3 ☐ 4 to 6 years
4 ☐ 7 to 8 years
5 ☐ 9 to 11 years
6 ☐ High school graduate
7 ☐ Some college or college graduate

Please continue on next page

8. What is the usual occupation of the child's parent or guardian? *(For example, carpenters, laborers, etc.)*

Father or male guardian	Mother or female guardian
____________________	____________________

9. Are they now employed?

Father or male guardian	Mother or female guardian
1 ☐ Now employed	1 ☐ Now employed
2 ☐ Now unemployed	2 ☐ Staying home (keeping house only)

10. How many months did they work during the past year? *(July 1, 1965 to June 30, 1966)*

Father or male guardian	Mother or female guardian
1 ☐ 12 months	1 ☐ 12 months
2 ☐ 10 or 11 months	2 ☐ 10 or 11 months
3 ☐ 7 to 9 months	3 ☐ 7 to 9 months
4 ☐ 2 to 6 months	4 ☐ 2 to 6 months
5 ☐ 1 month or less	5 ☐ 1 month or less
6 ☐ Did not work	6 ☐ Did not work

The following questions are about the home the child is staying in now.

11. Please mark an "X" in the box which tells where the house is located.

1 ☐ On a farm
2 ☐ In the country, but not on a farm
3 ☐ In a city or town
4 ☐ In the suburb or on the outskirts of a city or town

12. How many rooms in the house or apartment are regularly used for sleeping?

Number of rooms ____________

13. Is there running water inside the house?

1 ☐ Yes
2 ☐ No

The following questions are about the child who is now in the Head Start school and about any other children you listed in questions 4 or 5.

14. Has the child who is now in the Head Start school been in a Head Start, kindergarten, or nursery class before?

1 ☐ Yes – In a Head Start class
2 ☐ Yes – In kindergarten or nursery class
3 ☐ No

15. Have any of the other children been in a Head Start, kindergarten, or nursery class?

1 ☐ Yes – In a Head Start class
2 ☐ Yes – In kindergarten or nursery class
3 ☐ No

Please continue on next page

16a. When your child in the Head Start school was examined by the doctor or dentist was anything found to be the matter with him or her?

1 ☐ Yes
2 ☐ No
3 ☐ Not examined
4 ☐ Don't know

b. If something was found wrong, did the child get treatment for it?

1 ☐ Yes
2 ☐ No
3 ☐ Don't know

c. Where was treatment given?

1 ☐ Doctor's or dentist's office
2 ☐ Clinic
3 ☐ Hospital
4 ☐ Other *(Specify)* ____________________

The following questions are about the family with whom the child is living. They do not apply to any people you listed in question 6.

17. Does the family receive any public welfare?

1 ☐ Yes
2 ☐ No

18. In which of the following income groups would the child's family be?
(Count income of all family members. Include earnings, welfare, assistance, and all other kinds of income.)

1 ☐ Less than $1,000 a year
2 ☐ $1,000 to $1,999 a year
3 ☐ $2,000 to $2,999 a year
4 ☐ $3,000 to $3,999 a year
5 ☐ $4,000 to $4,999 a year
6 ☐ $5,000 to $5,999 a year
7 ☐ $6,000 to $7,999 a year
8 ☐ $8,000 to $9,999 a year
9 ☐ $10,000 or more a year

19. Please mark an "X" in the box which shows whether the family has the use of:

a. A car or truck . . . . 1 ☐ Yes 2 ☐ No
b. A radio . . . . 1 ☐ Yes 2 ☐ No
c. A television set . . . . 1 ☐ Yes 2 ☐ No
d. A telephone . . . . 1 ☐ Yes 2 ☐ No

20. Is there anyone in the family who usually gets a newspaper?

1 ☐ Yes 2 ☐ No

If "Yes" to question 20 — How often?

1 ☐ Every day
2 ☐ At least once a week
3 ☐ Less than once a week

Thank you very much for your cooperation. Please use this space if you would like to say something about the Head Start school program that may aid the Office of Economic Opportunity in providing better programs in the future. For example: the need for transportation, different hours, longer programs, etc.

____

EXHIBIT D-3 SPANISH VERSION OF FAMILY INFORMATION FORM (CAP-HS FORM 46a)

PRC R-886
D-26

OFFICE OF ECONOMIC
OPPORTUNITY

EXECUTIVE OFFICE OF THE PRESIDENT
WASHINGTON, D.C. 20506

BUDGET BUREAU NO. 116-R052
APPROVAL EXPIRES JUNE 30, 1967

Estimado Padre o encargado:

La Escuela "Head Start" en la cual está matriculado su niño, es parcialmente financiada por la Oficina de Oportunidades Económicas en Washington, D.C. Para poder explicar el programa al Congreso de los Estados Unidos y al público en general, la Oficina de Oportunidades Económicas necesita ciertos datos acerca de las familias que se benefician de este programa.

Las respuestas que usted nos dé a las preguntas aquí incluídas serán combinadas con miles de otras que vendrán de todas partes de los Estados Unidos. Estas respuestas nos darán información acerca de las familias que están matriculando sus niños en el Programa "Head Start."

Si usted tiene algún problema al contestar las preguntas, favor de pedir ayuda al Centro de Desarrollo del Niño "Head Start." Cuando termine de contestar esta forma, favor de enviarla inmediatamente por correo usando el sobre que le incluímos, el cual no necesita sello.

Atentamente,

Jule M. Sugarman

Jule M. Sugarman
Director Asociado
Project Head Start

Si no se le incluyó un sobre,
favor de enviar esta forma
por correo a:

Project Head Start (Summer)
Census Operations Office
1201 East Tenth Street
Jeffersonville, Indiana 47130

CAP-HS FORM 46a (6-21-66)

Las siguientes preguntas son acerca de las personas que regularmente viven con el niño matriculado en la escuela "Head Start." Conteste estas preguntas marcando una "X" en el encasillado correcto, o anote un número en el espacio provisto.

LISTA DE PERSONAS		NUMERO Y EDADES DE PERSONAS
1. ¿Hay un padre, padrastro o padre de crianza viviendo con el niño?	1 ☐ Sí → 2 ☐ No	**1a. ¿Qué edad tiene?** 1 ☐ Menor de 21 años 2 ☐ 21-54 años 3 ☐ 55-64 años 4 ☐ 65 años o más
2. ¿Hay una madre, madrastra o madre de crianza viviendo con el niño?	1 ☐ Sí → 2 ☐ No	**2a. ¿Qué edad tiene?** 1 ☐ Menor de 21 años 2 ☐ 21-54 años 3 ☐ 55-64 años 4 ☐ 65 años o más
SI NO HAY PADRE O MADRE EN EL HOGAR: **3. ¿Quién es el encargado del niño (la persona mayormente responsable del cuidado del niño)?**	1 ☐ Un hermano, tío, abuelo u otro familiar varón → 2 ☐ Una hermana, tía, abuela u otro familiar mujer → 3 ☐ Un hombre no relacionado → 4 ☐ Una mujer no relacionada →	**3a. ¿Qué edad tiene la persona encargada?** 1 ☐ Menor de 21 años 2 ☐ 21-54 años 3 ☐ 55-64 años 4 ☐ 65 años o más
4. ¿Tiene el niño hermanos o hermanas viviendo en el hogar? *(No incluya el hermano o hermana que actúa como persona encargada.)*	1 ☐ Sí → 2 ☐ No	**4a. ¿Cuántos hay?** — Número Menores de 6 años . . . . ☐ 6-15 años . . . . ☐ 16-21 años . . . . ☐ 22 años o más . . . . ☐
5. ¿Tiene el niño otros parientes (tales como abuelos, primos, etc.) que viven en el hogar con él? *(No incluya el pariente que actúa como encargado del niño.)*	1 ☐ Sí → 2 ☐ No	**5a. ¿Cuántos hay?** — Número Menores de 6 años . . . . ☐ 6-15 años . . . . ☐ 16-21 años . . . . ☐ 22-54 años . . . . ☐ 55-64 años . . . . ☐ 65 años o más . . . . ☐
6. ¿Existe alguna otra persona que vive en la misma casa pero que no es pariente del niño? *(No incluya la persona que actúa como encargada del niño.)*	1 ☐ Sí → 2 ☐ No	**6a. ¿Cuántas personas viven en la misma casa con el niño que no son familiares del niño?** Número ____________

Si el niño vive con sus padres, las siguientes preguntas son acerca de ellos.

Si el niño no vive con sus padres, las siguientes preguntas son acerca de su encargado.

7. Favor de marcar con una "X" en el encasillado que indica el grado de escuela mas alto que los padres o encargado cursaron.

Padre o engargado:

1 ☐ No asistió a la escuela
2 ☐ 1 a 3 años
3 ☐ 4 a 6 años
4 ☐ 7 a 8 años
5 ☐ 9 a 11 años
6 ☐ Graduado de Escuela Superior
7 ☐ Asistió a la universidad o graduado de universidad

Madre o encargada:

1 ☐ No asistió a la escuela
2 ☐ 1 a 3 años
3 ☐ 4 a 6 años
4 ☐ 7 a 8 años
5 ☐ 9 a 11 años
6 ☐ Graduado de Escuela Superiod
7 ☐ Asistió a la universidad o graduado de universidad

Favor de continuar en la página siguiente

8. ¿Cuál es la ocupación acostumbrada del padre o encargado del niño? *(Por ejemplo, carpintero, abrero, etc.)*

Padre o encargado	Madre o encargada
____________________	____________________

9. ¿Están empleados actualmente?

Padre o encargado:	Madre o encargada:
1 ☐ Empleado actualmente	1 ☐ Empleada actualmente
2 ☐ Desempleado actualmente	2 ☐ En la casa *(Cuidano del hogar solamente)*

10. ¿Cuántos meses trabajaron durante el año pasado? *(Julio 1, 1965 a Junio 30, 1966)*

Padre o encargado:		Madre o encargada:	
1 ☐ 12 meses	4 ☐ 2 o 6 meses	1 ☐ 12 meses	4 ☐ 2 o 6 meses
2 ☐ 10 o 11 meses	5 ☐ 1 mes o menos	2 ☐ 10 o 11 meses	5 ☐ 1 mes o menos
3 ☐ 7 o 9 meses	6 ☐ No trabajó	3 ☐ 7 o 9 meses	6 ☐ No trabajó

Las siguientes preguntas son acerca del hogar en que el niño vive actualmente.

11. Favor de marcar con una "X" en el encasillado que indíca donde está localizada la casa

1 ☐ En una finca, granja o rancho
2 ☐ En el campo pero no en una finca, granja o rancho
3 ☐ En la ciudad o pueblo
4 ☐ Suburbio, o afuera de la ciudad o pueblo

12. ¿Cuántos cuartos en la casa o apartamento se usan regularmente para dormir?

Número de cuartos ____________________

13. ¿Hay agua corriente en la casa?

1 ☐ Sí 2 ☐ No

Las siguientes preguntas son acerca del niño que está en la escuela "Head Start" y todos los otros niños mencionados en las preguntas 4 y 5.

14. ¿Asistió el niño que está actualmente en la escuela "Head Start" anteriormente a otra clase "Head Start," a una escuela de niños (Kindergarten), o a una escuela maternal (Nursery)?

1 ☐ Sí – A clase de "Head Start"
2 ☐ Sí – A escuela de niños (Kindergarten) o escuela maternal (Nursery)
3 ☐ No

15. ¿Ha asistido algún otro de los niños en su casa a escuela "Head Start", escuela de niños (Kindergarten), o escuela maternal (Nursery)?

1 ☐ Sí – A clase de "Head Start"
2 ☐ Sí – A escuela de niños (Kindergarten) o escuela maternal (Nursery)
3 ☐ No

Favor de continuar en la página siguiente

16a. Cuando su niño fue examinado por el doctor o dentista en la Escuela Headstart, ¿se le recomendó a usted que el niño recibiera algun tratamiento?

1 ☐ Sí
2 ☐ No
3 ☐ No se examinó
4 ☐ No sé

b. Si algún tratamiento fué recomendado, ¿recibió el niño este tratamiento?

1 ☐ Sí
2 ☐ No
3 ☐ No sé

c. Si recibió el tratamiento, ¿En dónde fue este tratamiento dado?

1 ☐ Oficina del Doctor o Dentista
2 ☐ Clínica Privada
3 ☐ Hospital
4 ☐ Otro sitio ____________________

Las siguientes preguntas son acerca de la familia con quien al niño vive. No se aplican a ninguna de las personas incluídas en la pregunta 6.

17. ¿Recibe la familia ayuda de Bienestar Público?

1 ☐ Sí
2 ☐ No

18. ¿A cuál de los siguientes grupos de entrada anual pertenece la familia del niño? *(Considere el sueldo de todos los miembros de la familia que viven en el hogar. Incluya dinero recivido en sueldo, ayuda pública, asistencia, etc.)*

1 ☐ Menos de $1,000 al año
2 ☐ $1,000 a $1,999 anual
3 ☐ $2,000 a $2,999 anual
4 ☐ $3,000 a $3,999 anual
5 ☐ $4,000 a $4,999 anual
6 ☐ $5,000 a $5,999 anual
7 ☐ $6,000 a $7,999 anual
8 ☐ $8,000 a $9,999 anual
9 ☐ $10,000 o más al año

19. Favor de marcar con una "X" en el encasillado que indica si la familia usa:

a. Automovil o camión	1 ☐ Sí	2 ☐ No
b. Radio	1 ☐ Sí	2 ☐ No
c. Televisión	1 ☐ Sí	2 ☐ No
d. Teléfono	1 ☐ Sí	2 ☐ No

20. ¿Hay alguna persona en la familia que adquiere el periódico regularmente?

1 ☐ Sí   2 ☐ No

↓

Si la contestación es "Sí" a la pregunta 20 – **¿Cuán a menudo?**

1 ☐ Todos los días
2 ☐ Por lo menos una vez a la semana
3 ☐ Menos de una vez a la semana

Muchas gracias por su cooperación. Por favor use este espacio si quiere decir algo acerca del programa de escuelas "Head Start" que pueda ayudar a la Oficina de Oportunidades Economicas a proveer mejores programas en el futuro. Por ejemplo: la necesidad de transportación, horas diferentes, programas mas largos, etc.

____

EXHIBIT D-4 BEHAVIOR INVENTORY (CAP-HS FORM 37)

PRC R-886
D-31

Budget Bureau No. 116-R055; Approval Expires June 30, 1966

CAP-HS Form 37
(5-4-66)

OFFICE OF ECONOMIC OPPORTUNITY
PROJECT HEAD START

BEHAVIOR INVENTORY

Child's name			School		
Grant No.	Center No.	Class No.	Child No.	Examiner's identification	Date

Present week of center's operation	**INSTRUCTIONS** Please indicate as accurately as possible how this child behaves by marking one of the four responses to each question. Base your response to every item on your personal observation and experience with the child.

	Very much like	Some-what like	Very little like	Not at all like
	1	2	3	4
1. Is usually carefree; rarely becomes frightened or apprehensive				
2. Is sympathetic, considerate, and thoughtful toward others				
3. Is easily distracted by things going on around him				
4. Is very suggestible; lets other children boss him around				
5. Talks eagerly to adults about his own experiences and what he thinks				
6. Is unduly upset or discouraged if he makes a mistake or does not perform well				
7. Often keeps aloof from others because he is uninterested, suspicious, or bashful				
8. Defends or praises his own efforts				
9. Is confident that he can do what is expected of him				
10. Is jealous; quick to notice and react negatively to kindness and attention bestowed upon other children				
11. Is methodical and careful in the tasks that he undertakes				
12. Is rarely able to influence other children by his activities or interests				
13. Tries to figure out things for himself before asking adults or other children for help				
14. Greatly prefers the habitual and familiar to the novel and the unfamiliar				
15. Appears to trust in his own abilities				
16. Has little respect for the rights of other children; refuses to wait his turn, usurps toys other children are playing with, etc.				
17. Seems disinterested in the general quality of his performance				
18. Responds to frustration or disappointment by becoming aggressive or enraged				
19. Is excessive in seeking the attention of adults				
20. Sticks with a job until it is finished				
21. Goes about his activities with a minimum of assistance from others				
22. Is constricted, inhibited, or timid; needs to be urged before engaging in activities				
23. Is even-tempered, imperturbable; is rarely annoyed or cross				
24. Is reluctant to talk to adults; responds verbally only when urged				
25. Works earnestly at his classwork or play; does not take it lightly				
26. Is often quarrelsome with classmates for minor reasons				

Please continue on reverse side

	Very much like	Some-what like	Very little like	Not at all like
	1	2	3	4
27. Does not need attention or approval from adults to sustain him in his work or play				
28. When faced with a difficult task, he either does not attempt it or gives up very quickly				
29. Does not like to be interrupted when engaged in demanding activities, e.g., puzzles, painting, constructing things				
30. Welcomes changes and new situations; is venturesome, explores, and generally enjoys novelty				
31. Calmly settles difficulties that arise without appeal to adults or others				
32. Is reluctant to use imagination; tends not to enjoy "make-believe" games				
33. Likes to talk with or socialize with the teacher				
34. Often will not engage in activities unless strongly encouraged				
35. Is eager to inform other children of the experiences he has had				
36. Emotional response is customarily very strong; over-responds to usual classroom problems, frustrations, and difficulties				
37. Is uncooperative in group activities				
38. Is usually polite to adults; says "Please," "Thank you," etc.				
39. Asks many questions for information about things, persons, etc. (*Emphasis here should be on questions prompted by genuine curiosity rather than bids for attention.*)				
40. Usually does what adults ask him to do				
41. Requires the company of other children; finds it difficult to work or play by himself				
42. Responds to frustration or disappointment by becoming sullen, withdrawn, or sulky				
43. Demonstrates imaginativeness and creativity in his use of toys and play materials				
44. Insists on maintaining his rights, e.g., will not yield his place at painting, or at the carpentry bench, etc.; insists on getting his turn on the slide or in group games, etc.				
45. Is wanted as a playmate by other children				
46. Is lethargic or apathetic; has little energy or drive				
47. Has a tendency to discontinue activities after exerting a minimum of effort				
48. Is generally a happy child				
49. Approaches new tasks timidly and without assurance; shrinks from trying new things				
50. What he does is often imitated by other children				

DO NOT MARK IN THIS SPACE

APPENDIX E

EXAMINATIONS OF SELECTED CENTERS

A. Introduction

Although the major thrust of the Head Start study was not directed at comparative evaluation of specific staff, facilities, or center programs, an effort has been made to identify, ex post facto, some of the most and least effective Child Development Centers.

What are the elements which might mark a particular center as more or less effective? Obviously, there were great variations from center to center in terms of staff, physical plant, programing, and children served, not to mention intercenter "philosophical" differences.

Staff varied with respect to such factors as race, age, level of education, and amount and type of experience with young children. The staff of the Child Development Centers, then, could not be labeled as homogeneous.

The centers varied in size, layout of rooms, furnishings, equipment, lighting, ventilation, outdoor space available, etc. Head Start classes were held in facilities ranging from damp, dim church basements to model air-conditioned classrooms with wall-to-wall carpeting.

In some centers there was considerable emphasis placed upon cognitive stimuation and language development, while in other centers staff efforts seemed to be aimed more at meeting certain social and emotional needs of the children. While general guidelines for programs were provided by Head Start headquarters, daily schedules for individual centers varied greatly regarding routines, amount and type of structured activity, free play, etc. Where there was more than one class in a center, intracenter differences were often marked. Standardization of programing was certainly not a characteristic of Head Start classes.

Finally, the children varied in ethnic and socio-economic background, age, achievement, attendance, and other measures. A 3-year-old, Spanish-speaking girl of PPVT IQ 70, from a "broken home" in a

Los Angeles slum is indeed different from a 6-year-old, English-speaking boy of PPVT IQ 100 from an intact home in a small midwestern community. Yet, both could have been Head Starters.

In summary, one cannot help but be impressed by the sheer diversity of the personnel, plants, programs, and enrollees in the total Head Start operation.

Efforts have been made before to find the factors that distinguish "successful" or good CDC's from poor ones.[1] The problem, methodologically as well as conceptually, is complex (see Section V). In this appendix, three approaches to the examination of center or direct program variables are presented. The first consists of identifying high and low ranking centers on the basis of raw score indices and examining associated information. One form of this approach is to use an age-corrected index, such as the center average PPVT IQ. A second index is to use the center residuals--that is, the deviations of the center means from their fitted values after regression. Again, the highs and lows can be singled out and examined. A third approach, admittedly weak, given the design of the present study, is to compare classes within centers in an attempt to identify outstanding classes.

The three methods are presented here more as demonstrations than as rigorous and planned primary investigations. Again it should be recalled that this study did not utilize pre-testing; it deals with comparisons of end test standings, not with gain scores.

B. Raw Score (RS) and IQ Indices

In order to obtain at least a crude ranking of individual centers as a function of children's performance, an unweighted average of raw score means for each center was obtained by:

$$\begin{matrix}\text{PPVT RS}\\\text{center mean}\end{matrix} + \begin{matrix}\text{BI RS}\\\text{center mean}\end{matrix} + \begin{matrix}\text{PSI RS}\\\text{center mean}\end{matrix} + \begin{matrix}\text{VSMS RS}\\\text{center mean}\end{matrix}$$

---

[1] See, e.g., Ozer, Mark. The Effects of Neurological and Environmental Factors on the Language Development of Head Start Children: An Evaluation of the Head Start Program. Report of OEO Contract 528, 1965.

These sums of raw score means were ranked from highest to lowest. The top ten centers (those with the ten highest sums of raw score means) were compared with the lowest ten centers. A comparison of the former (hereafter referred to as the TT centers) with the latter (BT centers), reveals a number of interesting trends (see Exhibit E-1).

1. Race

Judging from the composition of the groups tested, the TT centers served a largely white population, while the BT centers served a largely non-white population. Although the number of male and female children tested at the two sets of centers corresponded rather closely, the racial composition was essentially reversed.

Five of the TT centers had no Negro children tested, and none of the remaining five had more than five Negroes tested. Six of the BT centers had exclusively Negro enrollees. In the TT centers, 87-1/2 percent of the children tested were white. In the BT centers, 88 percent of the children were non-white.

2. Age

The mean age of the 84 males and 52 females tested in the TT centers was 71.5 months (or nearly 6 years). In contrast, the mean age of the 82 males and 57 females tested in the BT centers was 55.1 months (or approximately 4-1/2 years). Thus, the tested population of the TT centers was, on the average, 16.4 months older than the BT center children tested.

In eight of the 10 TT centers, the mean age of the children tested was greater than 70 months, while in none of the BT centers did the mean age reach 70 months. In fact, in all but one BT center it was less than 60 months.

3. Degree of Urbanization

Four of the TT centers were classified as rural, two as suburban, and four as urban. However, all 10 of the BT centers were urban. Interestingly, five of these were located in the same city.

EXHIBIT E-1 SELECTED CHARACTERISTICS OF THOSE CENTERS WITH THE TOP TEN AND THE BOTTOM TEN SUMS OF RAW SCORE MEANS

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
				Number of Children Tested										
Center Number	Rank of Sum of Raw Score Means	Center Type and Center Average Duration in Weeks	Urbanization	Male White	Female White	Male Non-White	Female Non-White	Mean Age	Mean Siblings	Number of Classes	Number of Children in Each Center	Attrition Rate	Mean Size of Classes	Approximate Total Hours in Program
28	1	M 22.0	Suburban	8	4	-	-	73.08	-	2	22	0.88	11.0	275.0
07	2	S 13.0	Rural	5	4	2	3	71.77	-	3	32	0.71	10.7	325.0
45	3	M 23.0	Suburban	5	3	2	2	73.66	5.8	2	34	0.93	17.0	402.5
42	4	M 19.0	Rural	9	5	-	1	73.73	4.2	2	38	0.97	19.0	380.0
41	5	M 19.0	Rural	4	5	-	2	74.09	3.9	1	15	0.60	15.0	332.5
51	6	M 14.9	Rural	12	3	-	-	61.46	3.5	1	15	0.88	15.0	280.0
36	7	M 18.6	Urban	6	9	-	-	75.33	4.1	1	21	0.94	21.0	585.0
01	8	S 8.7	Urban	11	3	-	-	76.21	3.8	3	54	0.94	18.0	100.0
06	9	S 10.0	Urban	11	3	-	-	74.86	4.0	5	100	0.98	20.0	150.0
05	10	S 11.4	Urban	5	4	4	1	60.43	3.1	2	29	0.94	14.5	165.0
Subtotal		(mean) 15.96		76	43	8	9	(mean) 71.50	(mean) 4.1	(mean) 2.2	(mean) 36.0	(mean) 0.877	(mean) 16.1	(mean) 299.5
Total				119		17								
59	62	L 29.0	Urban	-	-	7	6	57.38	2.8	4	97	0.82	24.3	290.0
33	63	M 16.6	Urban	1	-	9	4	51.00	3.1	3	62	0.90	20.7	440.0
13	64	S 11.5	Urban	-	-	6	8	54.40	6.8	4	60	0.90	15.0	165.0
68	65	L 24.4	Urban	-	-	9	6	52.33	2.8	2	27	0.62	13.5	720.0
02	66	S 10.7	Urban	-	2	10	1	60.90	5.4	4	62	0.94	15.5	133.7
16	67	S 13.9	Urban	-	-	8	6	57.50	5.1	2	30	1.00	15.0	195.0
44	68	M 18.1	Urban	-	-	7	6	53.92	3.1	3	32	0.58	10.7	540.0
38	69	M 17.6	Urban	4	2	4	5	53.26	2.4	2	22	0.62	11.0	510.0
69	70	L 25.6	Urban	4	4	4	2	56.29	6.3	3	55	0.30	18.3	750.0
43	71	M 20.4	Urban	-	-	9	5	53.86	3.2	2	35	1.00	17.5	600.0
Subtotal		(mean) 19.5		9	8	73	49	(mean) 55.10	(mean) 4.1	(mean) 2.9	(mean) 48.2	(mean) 0.778	(mean) 16.2	(mean) 434.4
Total				17		122								

4. Number of Classes

The TT centers tended to have, on the average, a somewhat smaller number of classes (2.2) as compared with the BT centers (2.9). Three of the TT centers had only one class, while none of the BT centers had only one class.

5. Class Size

The mean size of individual classes in the TT and BT centers was essentially the same--16.1 and 16.2, respectively. There was a considerable range of class sizes, however. In the TT centers, the mean class size ranged from 10.7 to 21 children, whereas in the BT centers it ranged from 10.7 to 24.3 children.

6. Size of Centers

The TT centers averaged fewer children enrolled "in the center" (36) as compared with the BT centers (48.2). With respect to population, then, the BT centers were 25 percent larger than the TT centers. (This trend could be expected, since it has already been noted that the BT centers had a larger mean number of classes.)

7. Length of Program

Among the TT centers, there were four S, six M, and no L centers, while for the BT centers the corresponding figures were three, four, and three, respectively.

The mean length of program in weeks for the TT centers was approximately 16, while for the BT centers the figure was 19.5. Thus, the BT centers spanned, on the average, 3.5 more weeks than the TT centers. However, it should be remembered that the designation of a center as of the S, M, or L type may be misleading if the total number of hours of the program offered at a particular center is not taken into account. With the great variation in daily length of program it was possible, for example, for children in certain S centers to be exposed to more hours of program than children in some M centers or even in an L center. For example, children who attended one TT center--an S center of 13 weeks' duration--were offered 325 hours of program.

Yet children who attended one of the BT centers--an L center of 29 weeks' duration--were offered only 290 hours of program.

Therefore, due to the great variation in daily length of program, the total number of hours of program by center was computed for the TT and BT centers. This figure ranged from a low of 100 hours (a TT center) to a high of 750 hours (a BT center).

The mean number of hours of program was 299.5 for the TT centers and 434.4 for the BT centers.

Only one TT center offered more than 500 hours of program, compared with five BT centers which did. However, it should be noted that at least one of the TT centers gave a relatively large "dose" of program (585 hours), while at least two BT centers gave relatively small "doses" of program (165 and 195 hours). Stated somewhat differently, the TT centers operated, on the average, for 31 percent less time than the BT centers.

The exhibits found in this section should be read with the understanding that figures dealing with class size and other aspects of pupil attendance are often necessarily only approximations. The enrollment situation in the Head Start Child Development Centers ranged from firm to fluid. In some centers, the pupil accounting was concise and current, while in other centers attendance records were more loosely kept, and day-to-day enrollment figures were open to various interpretations.

To give some idea of the fluidity of the "population picture" in a specific center, let us examine excerpts from the notes of a tester who tested a center in a large Eastern city. These excerpts have been selected from six handwritten pages of one tester's notes and are quoted directly.

> Attendance at this center is very sporadic. Some of the children come once a week, some come twice. Many of the children who are still enrolled really do not come. It should be noted that because of the sporadic attendance many of the children really have not been going for 20 weeks.

> On the books, 36 children are enrolled; only 3 of the original 35 have dropped out. However, while this tester was at the center there were certainly not 36 children! Out of the 25 numbers given me, 8 of them definitely would not be tested because the teachers said they really weren't coming (and they really weren't). It was rather difficult to get even a rough estimate of the number attending; however, in class 01 about 13 (of the 20 enrolled) usually came. Of course, this isn't necessarily the same 13 every day but 13 is about the number attending.
>
> In class 02 about 10 to 12 children regularly attend. Incidentally, 'regular' here involves a value judgment. A couple times a week is 'regularly,' mainly because there are a lot who don't come at all.

In summary, when all the Head Start Child Development Centers were ranked in order of "sum of raw score means," the bottom ten centers (as compared with the top ten) tended to be more "big city" and more Eastern geographically, of longer duration in terms of both number of weeks spanned and hours of program, and larger in terms of both number of classes and total population enrolled.

Tested children of the BT centers tended to be non-white and younger than those in TT centers. Tested children of both the BT and TT centers had a similar mean number of siblings and were enrolled in classes of approximately the same mean size.

With 1.0 representing no attrition, the TT centers (with an attrition rate of 0.877) displayed a somewhat greater holding power than the BT centers (with an attrition rate of 0.778).

The TT centers tended to be less urbanized and more Western and Southwestern geographically. On the average, they lasted fewer weeks, offered fewer hours of program, and contained a somewhat smaller number of classes and fewer enrollees. Children were more often white and older.

8. Another Ranking

Exhibits E-2 and E-3 show those centers with the 10 highest and 10 lowest PPVT IQ means, respectively. Data from tester reports have been charted for these two groups of centers.

EXHIBIT E-2 DESCRIPTION OF CENTERS WITH THE TEN HIGHEST PPVT IQ MEANS

1	2	3	4	5	6	7	8	9		10		11			12			13		14		15
Center Number	Mean PPVT IQ	Rank of Mean PPVT IQ	Total Children Registered at Start	Total Registered and Attending at Time of Testing	No. of Originally Registered Children Who Completed Program	No. of Classes per Center	No. of Children in Each Class	Sampling Procedure Varied From Routine		Quiet, Uninterrupted Testing Location		Lighting			Heating/Ventilation			Easy to Establish Rapport With Children		Difficulties With Teacher Experienced by Tester		Mean Age in Months
								Yes	No	Yes	No	Poor	Fair	Adequate	Poor	Fair	Adequate	Yes	No	Yes	No	
51(1)	102.67	1	17	15	15	1	15		X	X				X			X	X			X	61.46
34	97.81	2	8	12	7	2	9, 5		X	X				X			X	X			X	61.36
28(1)	96.67	3	25	22	22	2	11		X	X				X			X	X			X	73.08
05(1)	96.07	4	31	29	29	2	14, 15		X	X				X		X		X		X(2)		60.43
70	94.43	5	41	37	38	3	20, 15, 15	X		X								X			X	54.14
39	94.27	6	18	16	16	1	16		X	X				X			X	X			X	61.08
41(1)	93.91	7	20	15	12	1	15		X	X				X			X	X			X	74.09
57	93.73	8	60	58	58	5	16		X	X				X			X	X			X	60.73
47	92.36	9	209	137	135	8	16 to 18		X	X				X			X	X	X		X	61.13
40	91.77	10	18	18	17	1	18		X	X				X			X	X			X	61.07
Totals			447	359	349	26		1	9	10				9		1	8	10	1	1	9	
Mean	95.46																					62.86

Notes: (1) These centers are also among those centers with the top ten sums of raw score means.
(2) Only minor difficulties were reported.

EXHIBIT E-3 DESCRIPTION OF CENTERS WITH THE TEN LOWEST PPVT IQ MEANS

1	2	3	4	5	6	7	8	9		10		11			12			13		14		15
Center Number	Mean PPVT IQ	Rank of Mean PPVT IQ	Total Children Registered at Start	Total Registered and Attending at Time of Testing	No. of Originally Registered Children Who Completed Program	No. of Classes per Center	No. of Children in Each Class	Sampling Procedure Varied From Routine		Quiet, Uninterrupted Testing Location		Lighting			Heating/Ventilation			Easy to Establish Rapport With Children		Difficulties With Teacher Experienced by Tester		Mean Age in Months
								Yes	No	Yes	No	Poor	Fair	Adequate	Poor	Fair	Adequate	Yes	No	Yes	No	
38(2)	73.07	61	35	25	22(3)	2	20,16	X		X			X			X		X			X	53.26
27	72.93	62	60	57	60(1)	4	15	X		X				X			X	X			X	54.66
04	71.79	63	14	15	14	1	15		X	X				X			X	X			X	74.35
72	71.62	64	30	40	15	3	15,20,17	X		X				X			X	X			X	57.92
68(2)	69.87	65	35(3)	27	22	2	20,15		X		X		X		X			X			X	52.33
08	68.77	66	167	141	141	10	15	X		X				X			X		X		X	71.77
03	68.61	67	36	47	33	3	15,16,16		X	X				X			X		X		X	75.92
69(2)	64.54	68	10	55	Unknown	3	14,21,20	X		X			X		X			X(5)			X	56.29
43(2)	64.14	69	15	35	15	2	20,15		X	X				X			X	X			X	53.86
23	61.00	70	57	57	47	4	14,28,28(4)	X		X				X			X	X			X	69.00
Totals			459	499		34		6	4	9	1		3	7	2	1	7	8	2		10	
Mean	68.63																					61.94

Notes: (1) Expected.
(2) These centers are also among those centers with the bottom ten sums of raw score means.
(3) Uncertain.
(4) One class (03) was split to form two classes.
(5) With certain qualifications.

Only four of the "top ten PPVT IQ mean centers" are also among those ten centers with the top ten sums of raw score means (PPVT plus BI plus PSI plus VSMS). Likewise, only four of the "bottom ten PPVT IQ mean centers" are among those ten centers with the bottom ten sums of raw score means.

Different criteria for center ranking, then, result in different groupings of centers. And there are many possible criteria for ranking of centers. PRC has used but two--sums of raw score means, and PPVT IQ means.

There is a wide range of PPVT IQ center means shown in Exhibits E-2 and E-3 (from a low of 61.00 to a high of 102.67). The mean of the PPVT IQ means is 68.63 for the bottom ten centers and 95.46 for the top ten centers. There is a difference, then, of nearly 27 mean PPVT IQ points between the top ten and bottom ten center PPVT IQ means.

This difference may have important implications for program content and teaching methodology. But it also may be yet further evidence of the differential test anxiety hypothesis discussed in Section V. In seven of the bottom ten PPVT IQ mean centers, the majority of the children tested were non-white, and in three of these seven the children tested were exclusively non-white. However, in one of the bottom ten centers, all of the children tested were white. In nine of the top ten PPVT IQ mean centers, the majority of the children tested were white, and five of these nine centers had samples which were exclusively white. However, in one of the top ten centers, all of the children tested were non-white. The mean age in months of the children tested in the top ten PPVT IQ mean centers was 62.86, as compared with 61.94 in the bottom ten.

Five of the bottom ten PPVT IQ mean centers were located in the same city. Two of the top ten centers were located in the same city, and one of the top ten centers was located in the same city which contained five of the lowest centers. This was the only top center with a totally Negro testing sample. Thirteen of the 14 children tested were male. Tester reports did not indicate a particularly outstanding program in evidence at this center. It is possible that these enrollees were simply less deprived to begin with. Their mean age of 54.14 months was the lowest of any of the top ten centers.

It can be seen that in six of the ten bottom centers, sampling procedures varied from the routine (i.e., original sample lists), while in only one of the top ten centers did such variation occur. The variation involved replacement sampling forced by absences or attendance irregularities of the children in original sampling lists.[1]

Ranking of the centers according to mean PPVT IQ results in a considerably different grouping of centers as compared with ranking of the centers according to the sum of raw score means on four tests. PPVT IQ center means ranged from "retarded" to "average."

In terms of tested children, all but one of the top ten PPVT IQ mean centers were predominantly white, while seven of the bottom ten PPVT IQ mean centers were non-white. However, again in terms of tested children, one top center was exclusively non-white and one bottom center was exclusively white. The children in the top ten centers were, on the average, approximately 1 month older than bottom center children.

Testers reported necessary variations from the routine[2] in more sampling procedures in the bottom ten centers.

9. Some Detailed Findings

Exhibit E-4 presents further selected test data of the TT and BT centers. Column 2 shows the sums of the raw score means of those ten centers which ranked highest on this criterion, and those ten centers which ranked lowest.

In column 4, it can be observed that none of the BT centers had a PPVT raw score mean equal to the lowest PPVT raw score mean of the TT centers. Stated differently, the lowest PPVT raw score mean of

---

[1] See Exhibits 5 and 6 for further description of untested children.

[2] That is, more requirements to use the alternative procedures supplied for such purposes.

EXHIBIT E-4 SELECTED TEST DATA OF THOSE CENTERS WITH THE TOP TEN AND THE BOTTOM TEN SUMS OF RAW SCORE MEANS

1	2	3	4	5	6	7	8	9	10	11	12	13
Center Number	Sum of Raw Score Means	Rank of Sum of Raw Score Means	PPVT Raw Score Mean	Rank PPVT	VSMS Raw Score Mean	Rank VSMS	PSI Raw Score Mean	PSI Rank	BI Raw Score Mean	Rank BI	Mean VSMS Social Quotient	Mean PPVT IQ
28	328.12	1	53.83	1	62.54	5	67.33	2	144.42	27	106.23	96.67
07	323.11	2	48.79	7	60.32	12	66.43	3	147.57	19	98.73	85.93
45	318.21	3	47.25	9	60.71	9	60.50	6	149.75	13	98.40	82.42
42	317.20	4	51.07	3	61.53	8	61.67	4	142.93	32	100.85	88.87
41	316.50	5	53.73	2	62.73	4	71.59	1	128.45	68	103.47	93.91
51	314.10	6	50.87	4	61.83	7	58.80	8	142.60	36	122.45	102.67
36	311.56	7	49.93	5	62.23	6	61.07	5	138.33	49	101.59	85.60
01	311.28	8	47.14	10	60.43	11	52.21	17	151.50	8	93.21	78.86
06	309.70	9	42.50	21	60.13	15	53.07	15	154.00	6	94.08	77.27
05	308.72	10	46.29	12	56.50	42	55.00	12	150.93	9	102.97	96.07
Mean	315.85		49.14		60.90		60.77		145.05		102.20	88.82
59	264.68	62	36.61	48	54.92	54	40.38	51	132.77	63	101.20	78.92
33	263.79	63	34.00	59	52.07	69	39.29	54	138.43	48	120.82	82.14
13	261.93	64	36.43	49	50.14	72	38.86	56	136.50	58	93.82	83.00
68	261.71	65	27.20	67	50.40	71	33.21	64	150.90	10	95.11	69.87
02	261.66	66	34.85	57	54.58	57	40.92	49	131.31	64	95.88	74.38
16	258.75	67	33.79	60	57.68	31	41.07	48	126.21	70	113.50	73.93
44	256.32	68	32.08	64	56.62	39	41.54	44	126.08	71	114.82	74.85
38	253.00	69	30.20	66	53.00	65	31.53	66	138.27	50	103.81	73.07
69	251.95	70	26.51	68	55.93	46	30.00	68	139.51	45	109.95	64.54
43	236.22	71	23.71	69	53.36	64	31.36	67	127.79	69	102.26	64.14
Mean	257.00		31.54		53.87		36.82		134.78		105.11	73.88

the TT centers was 42.5, while the highest PPVT raw score mean of the BT centers was 36.61. The mean of the PPVT raw score means for the TT centers was 17.6 points higher than for the BT centers. It can be seen from column 13 that the mean of the PPVT IQ means was 88.82 for the TT centers and 73.88 for the BT centers, a difference of nearly 15 mean IQ points. Four of the TT centers had PPVT IQ means above 90, while the highest PPVT IQ mean of any BT center was 83. The lowest PPVT IQ mean for any TT center was 77.27, while three of the BT centers had PPVT IQ means below 70. Seven of the ten TT PPVT IQ means were higher than any of the BT PPVT IQ means.

On the other hand, it is interesting to note (column 12) that the mean of the VSMS Social Quotient means for the BT centers was slightly higher than that for the TT centers. The children from the BT centers, then, had considerably lower PPVT IQ's, on the average, than the TT children, but averaged slightly higher VSMS Social Quotients.

Column 8 indicates that the highest PSI raw score mean for any BT center was 41.54, while the lowest PSI raw score mean for any TT center was 52.21. The mean of the PSI raw score means for the TT centers exceeded that for the BT centers by approximately 24 points.

The mean of the BI raw score means for the TT centers was approximately 10 points higher than for the BT centers. One BT center (68) had a BI raw score mean which was exceeded by only two TT centers, and one TT center (41) had a BI raw score mean which was exceeded by all but two BT centers; generally, however, TT BI raw score means exceeded those of the BT centers.

It can be seen that the TT center (41) with the low BI raw score mean had very high raw score means on the other three tests, while the BT center (68) with the high BI raw score mean had low raw score means on the other three tests. These findings are difficult to interpret.

Column 2 of Exhibit E-4 shows the sums of raw score means for the top ten and bottom ten centers. It must be stated that nearly 50 percent of the sums comes from the BI raw score means.

10. Summary

Those centers with the top ten sums of raw score means, as compared with those centers with the bottom ten sums of raw score means, exhibited the following characteristics: higher PFVT raw score means in all cases; a mean of PPVT raw score means 17.6 points higher; higher PPVT IQ means in seven of 10 cases; a mean of PPVT IQ means higher by nearly 15 points; no PPVT IQ center mean below 77.27; a slightly lower mean of VSMS Social Quotient means; a mean of PSI raw score means approximately 24 points higher; higher PSI raw score center means in all cases; and a mean of BI raw score means approximately 10 points higher.

C. Tester Informal Reports and Narratives

As part of their activity in the field, Head Start testers wrote informal reports and narratives concerning each Child Development Center visited. These materials comprise, in effect, brief descriptions of the testers' experiences at the site.

Since the major task of the testers was not a comprehensive evaluation of center facilities and functionings, they were not specifically trained for objective and extensive observation of centers and surroundings. However, perusal of tester informal reports and narratives can provide one with the anecdotal impressions of a particular center as seen by an "outsider"--a person not directly connected with the center operation.

Material concerning the previously identified "top ten" (TT) centers and "bottom ten" (BT) centers, in terms of sum of raw score means, has been extracted and organized under several headings (see Exhibits E-5 and E-6).

The testers' narratives did not follow a standard format, and there are gaps in the chart. The testers' wording has been shortened and, in some cases, paraphrased to fit into limited space.

1. Testing Conditions

It can be noted that testing conditions in four of the TT centers were described in positive terms. No negative testing conditions

EXHIBIT E-5 ANECDOTAL MATERIAL FROM TESTERS' REPORTS OF THOSE CENTERS WITH THE TOP TEN SUMS OF RAW SCORE MEANS

1 Center Number	2 Rank of Sum of Raw Score Means	3 Tests, Testing Conditions	4 Personnel	5 Community	6 Center (General)
28	1	Well-adjusted children	Very cooperative, willing to help	Residential suburb; growing commercial area	-
7	2	"Successful testing"	"On the whole, very cooperative"	Farm community of 1,500 people, many of Indian heritage	-
45	3	Very good	Very cooperative	"Rather large town, industrial center"	-
42	4	-	Very test-conscious; prior instruction of children geared toward tests	Population of from 900 to 1,000; rural community outside of a medium-sized city	Converted Army barracks
41	5	-	Very cooperative	Population of about 2,000; farming community	Attendance problem because of inadequate transportation
51	6	Excellent testing facilities	-	Wood processing plant is town's only industry. "Town does not appear run down;" there seem to be "many retired old folk, many on welfare."	-
36	7	Excellent testing site, almost completely free from disturbance	-	Population of 100,000; laboring class in quiet, residential area	Excellent meals; well-organized center with excellent, well-utilized facilities
01	8	-	"Hyperactive principal, needed constant reassurance as to plans, objectives, and impressions; extremely nervous and anxious to please."	-	-
06	9	-	Good teachers and aides	Country area with population of 355,000; fast-growing area	-
05	10	-	-	School located in midst of large government housing project in slum area; large number of Head Start children from fatherless homes, however great community interest expressed	-

EXHIBIT E-6 ANECDOTAL MATERIAL FROM TESTERS' REPORTS OF THOSE CENTERS WITH THE BOTTOM TEN SUMS OF RAW SCORE MEANS

1	2	3	4	5	6
Center Number	Rank of Sum of Raw Score Means	Tests, Testing Conditions	Personnel	Community	Center (General)
59	62	-	Could not seem to cope with children; because of disciplinary problem, they were unable to help them.	Row houses, high-rise apartments; as a whole, not a well-kept neighborhood. Low-income, semiskilled laboring families	"General impression very poor;" disciplinary problems with children interfered with success of classroom activities.
33	63	Good testing facilities	Very helpful, cooperative	Population about 560,000; many of Polish origin	"Facilities generally more than adequate;" located in a church; good nutrition
13	64	Poor arrangements, conditions	Two very cooperative; two indifferent with negative view of testing	-	-
68	65	Noise and interruptions, but this didn't seem to disturb the children	One commented that the teachers were not able to cope with children's problems. But another observation was that "the teachers spent time with the children and activities were directed."	Crowded, busy slum neighborhood; overpowering stench; gangs roaming streets; male vagrants abounded; high percentage of fatherless homes	Poor physical plant; "classes were often mass chaos."
02	66	-	-	City and environs (10-mile radius) have a population of 1,250,000. Population of the city, itself is 145,000. Retail center industrial area; many Spanish-speaking in area around center	-
16	67	Problems with whole staff resulting from hostility toward and general impatience with testers. (The center had experienced "continual mass intrusion and interference" through an earlier battery of tests.)	-	-	Noise, constant interruptions, general disorder
44	68	-	"Not overly friendly;" lack of insight into children's problems; ". . . seem to have created static, if frantic, situation."	Row house neighborhood. "Almost all parents are employed and living together."	Center located in dark, dank basement of church; high incidence of emotional problems among the children; seemingly adequate equipment
38	69	-	"Teachers were very poor. Their general concern was with the children's (comportment)." From lengthy description, teachers were untrained and inept.	Poor neighborhood; mixed ethnic groups of Italians, Polish, and Negroes; great idleness, vagrancy	Parents of children appeared quite young; sporadic attendance; slow children, poor responders
69	70	Distracting	Unprofessional; not aware of or committed to aims of Head Start	-	Nourishing meals
43	71	-	"No irritations or hostilities" in evidence; ". . . everyone was more than cooperative."	Downtown, semiresidential neighborhood; many parents on welfare, but most work; parental concern and interest high	"No outstanding medical, emotional, or intellectual problems" among the children who appeared "generally happy, curious, usually responsive and well-trained;" good atmosphere

were reported at these centers, while in the BT centers, negative testing conditions were reported in at least three cases and positive conditions in one.

2. Personnel

In five of the TT centers, personnel were described spontaneously as either cooperative or "good," with no negative impressions of TT center personnel recorded. The picture for the BT centers is considerably different. For at least one-half of these, testers had negative comments to make regarding the staff, and used such terms as "unprofessional," "very poor," and "indifferent." It should be added that in three of the BT centers, testers were positively impressed by staff cooperativeness. For several of the BT centers, testers made mention of noise and confusion. In several of the BT centers, maintenance of "discipline" seemed to be a major problem.

Resistance on the part of teachers to testing activity was noted in the two BT centers. Apparently this attitude stemmed from testing that had been done prior to arrival of the project tester.

3. Community

When the tester "community" comments for the TT centers are compared with those for the BT centers, it is found the latter centers were more often located in areas of a slum, "run-down" nature, with greater ethnic mix.

4. General

Positive comments regarding the nutritional component of the Head Start program were made by testers at three centers (one TT and two BT's). There were no negative comments concerning nutrition recorded.

5. Summary

In summary, testing conditions, personnel, learning environment, and neighborhood were reported as inadequate more often for the BT centers than for the TT centers.

The global impression of the BT centers derived from the tester reports is one of more noise and distraction and less effective staff--of more pupil behavior not under control. For example, at one BT center "Discipline problems with children interfered with the success of classroom activities," according to the tester. At a second BT center, "Classes were often mass chaos." At a third BT center, the tester reported "Noise, constant interruptions, and general disorder." No comments of this type were recorded for the TT centers.

Material from testers' informal reports and narratives concerning those centers with the top ten and bottom ten PPVT IQ means has been extracted and arranged under four headings in Exhibits E-7 and E-8. Blanks indicate that directly applicable material for a particular heading was not recorded by the tester, or was included in Exhibit E-5 or E-6.

D. Ranking Based on Center Residuals

The index of ranking in this method is, as described earlier, the center residual. Residuals were calculated for each dependent variable. Then the highest positive ten and the lowest negative ten centers, based on an unweighted composite or sum of residuals for the four tests, were located[1] and examined with respect to several types of information. The results are shown in Exhibit E-9. The centers involved are those with a check mark in the second column. They are given arbitrary numbers from 1 to 10 in the first column, and are starred if they were also in the top (or bottom) ten, based on the unadjusted composite index.

A composite index, consisting of the ten highest and lowest of the residuals for the PPVT center average raw score, was also used. The high and low centers thus identified were those with a check mark in the third column.

The results largely seem to speak for themselves. Among the bottom ten centers based on total composite residual value, six also had the lowest PPVT residual values. The average center size (with 3.4 classrooms) was slightly higher than that of the top ten centers.

---

[1] See Exhibit B-3 in Appendix B for a plot of PPVT center residuals against fitted values.

EXHIBIT E-7 ANECDOTAL MATERIAL FROM TESTERS' REPORTS OF THOSE CENTERS WITH THE TOP TEN PPVT IQ MEANS

1	2	3	4	5
Center Number	Community/Locale Description	Center Facilities	Children Tested, Testing Facilities, Situation	Teachers, Staff
51	-	-	-	-
34	This is a small town, 60 miles southwest of a major city. Industry includes coal mining and the production of paper tissues, cellulose wadding, asphalt, and roofing and insulating materials. In addition, the outlying areas are devoted to farming and the production of corn and other crops.	-	-	-
28	-	-	-	-
05	-	-	-	-
70	This is an entirely Negro area with several high-rise apartments nearby and several blocks of new duplex homes (well-kept). The surrounding area contains all poverty-type homes and slums.	Fairly new; a large community center building	Test consciousness in evidence on the part of teacher and child	Summary of tester's comments: Two of the teachers were generally not co-operative; a third was.
39	Located in the center of a popular resort area, this city has a population of 65,000. There are very few members of any minority group, and only a few Negroes.	-	-	-
41	-	-	-	-
57	-	-	-	-
47	This area appears to be semidivided into roughly three communities: one middle-class suburban area, one lower-middle-class rural area mainly consisting of whites, and a Mexican community that lives for the most part in post-World War II dwellings of very poor quality and little living space. Children in the program represented a pretty good cross-section of these communities. Nearly all children were brought to and from school by a school bus, with only a few brought by their parents. One or two lived close enough to walk.	Buildings specially constructed for the pre-school program and ideally suited for its needs Two sessions (a.m./p.m.) with classes grouped according to age	Adequate facilities; quiet, without distractions	Relationships with all the teachers were quite good, which made for a harmonious week.
40	Several of the worst tenements...though in need of repair, cleaning, and enlargement...appear palatial in comparison with homes of some other Head Start children. The great majority of the fathers of these (children) are employed, also in contrast with other Head Start children. On the whole, the community seems aware of its problems as well as assets and is attempting to take appropriate action.	Church in center of town, with children bussed to and from school	-	-

EXHIBIT E-8 ANECDOTAL MATERIAL FROM TESTERS' REPORTS OF THOSE CENTERS WITH THE BOTTOM TEN PPVT IQ MEANS

1	2	3	4	5
Center Number	Community/Locale Description	Center Facilities	Children Tested, Testing Facilities, Situation	Teachers, Staff
38	-	-	-	-
27	Small town, with a few small stores, for the most part residential and very quiet; surrounded by poor farming district. Some homes in the immediate area were shacks.	Located in a church which served as community center	-	-
04	Average-sized city of 125,000; suburban, mainly residential. Fishing and the military constitute the major source of livelihood.	Elementary school	Testing environment was above average. Children, although from deprived, welfare homes and of limited intelligence, were above average in taking care of their personal needs. Children were willing and eager.	Very cooperative
72	-	Adequate space/facilities; located in a church	Children were regular attendees.	Completely cooperative teachers, who were eager to do a good job, very affectionate toward the children, and enjoyed the center and the children very much
68	-	-	-	
08	Fruit, vegetable farm region; many migrant families; population of 15,000, the majority of Mexican extraction. Over 50 percent of the families were in the poverty category.	-	It was not easy to establish rapport with the children, who were painfully shy Mexican children not used to exposure to strangers.	In no way accountable for the children's attitudes; cooperative, very nice
03	-	-	Physical setting was very good; children were mostly Spanish-speaking children.	Cooperative, friendly; aides well-utilized
69	-	-	-	-
43	-	-	-	-
23	Population of 1,270, with 88 percent in the poverty category and population distribution of one-third each in white, Negro, and Spanish-American categories.	Located in elementary school	Children were extremely culturally deprived, and should be kept out of their own homes as long as possible.	Completely cooperative

EXHIBIT E-9 TOP TEN AND BOTTOM TEN CENTERS BASED ON RESIDUALS

Top Centers	Top Ten Based on Total Composite Residual Value	Top Ten Based on PPVT Residual	Program Length in Weeks	Proportion of Non-White	Center Size	Log Population	Region Number	Tester Number	Tester Comments
1*	X	X	16	0.00	1	2.915400	7	13	Excellent testing facilities; general cooperation
2	X	X	33	1.00	3	6.301575	2	15	Teacher and child test-conscious; two teachers cooperative, one uncooperative
3	X	X	11	1.00	3	6.394280	7	10	Adequate facilities; cooperative teachers; records incomplete and unorganized
4	X	X	11	0.29	2	5.677186	6	11	Testing facilities "fair;" teacher slightly uncooperative; records incomplete
5	X	X	27	1.00	2	6.550278	4	18	Definite interest in children evident
6	X	-	23	1.00	2	6.550278	4	14	Very good facilities; very cooperative teachers
7	X	-	23	0.00	8	3.444981	7	13	Adequate facilities
8*	X	X	22	0.00	2	3.942792	6	16	Well adjusted children; cooperative personnel
9*	X	-	13	0.36	3	3.020361	5	16	Successful testing; very cooperative personnel
10	X	X	15	0.00	2	3.624282	4	09	No special comment
11	-	X	15	0.38	2	4.096180	1	15	Adequate facilities; teachers cooperative
12	-	X	19	0.18	1	3.393575	6	16	Cooperative staff
13	-	X	30	1.00	2	6.550278	4	06	No special comment

EXHIBIT E-9 (Continued)

Bottom Centers	Based on Total Composite Residual Value	Based on PPVT Residual	Program Length in Weeks	Proportion of Non-White	Center Size	Log Population	Region Number	Tester Number	Tester Comments
1	X	-	19	0.00	1	4.956014	5	02	Well organized center; good facilities; full day program resulting in children seeming tired during afternoon session
2*	X	-	18	1.00	3	6.301575	2	14	Not overly friendly; lack of insight into children's problems; poor center facilities
3*	X	X	11	0.85	4	5.157345	1	15	Testing facilities adequate, but removed from actual center; one teacher very uncooperative
4	X	X	19	0.00	2	4.553303	7	13	Adequate facilities; good general cooperation
5*	X	X	18	0.60	2	6.301575	2	18	Poor, inept, and untrained teachers; poor attendance
6	X	X	13	0.00	10	4.148634	5	08	Very cooperative teachers; adequate testing facilities; children extremely passive
7*	X	-	14	1.00	2	6.394280	7	10	General disorder and constant interruptions at center; staff impatient and hostile to tester
8*	X	X	35	0.75	3	6.301575	2	05	Often distracted while testing; teachers unprofessional and not aware of Head Start aims
9	X	-	19	0.00	5	5.108953	6	04	Good staff cooperation
10*	X	X	20	1.00	2	6.301575	2	14	Cooperative staff; good atmosphere
11	-	X	22	0.60	4	6.550278	4	05	No special comment
12	-	X	24	1.00	2	6.301575	2	18	Very poor testing conditions; poor facilities; good cooperation
13	-	X	14	0.00	3	3.843357	5	02	Above average personnel; good cooperation; much parent involvement with staff
14	-	X	14	0.64	4	3.056905	5	03	Complete staff cooperation; children extremely culturally deprived

The number of all white or non-white centers was the same as the top ten, with a higher proportion of non-whites in the mixed centers. The testers' comments were favorable about eight of the centers, while six of the centers were cited for uncooperative or poor teachers and/or poor facilities.

No single significant factor or condition seems to stand out or to be attendant upon a center's high or low mean score. As has been pointed out, the absence of pre-testing has precluded any basis for comparison of relative gains or losses of the children during the course of the program.

E. Intracenter Class Differences

A third way of examining centers is to compare performance between classes within a center. The overall study was not specifically designed for this purpose. Nevertheless, sampling within centers tested was based on an approximately equal sample of children drawn from each class. In reality, of course, such was not always the exact result, owing to absences, etc.

To examine the effect of different classes within centers, all of the centers with two classes and all those with four classes in PRC's total sample were listed. Five centers from each group were then drawn at random. For each of these sample centers, a single classification analysis of variance was made for each of the four main tests (PPVT raw score, PSI total score, VSMS raw score, and BI total score). The results are given in Exhibit E-10. The exhibit shows, for each sample CDC analyzed, the number of classes, the tests used as dependent variables, the urbanization classification of the CDC, the type of CDC (S, M, or L), the percentage of non-white children in the CDC sample tested, the average age of the sample of children, and the total number of children in the sample. It also shows, for each dependent variable, the estimate of the treatment variance $\hat{\sigma}_A^2$ and of the child or error variance $\hat{\sigma}_e^2$, the ratio of between-mean squares to within-mean squares, and the F-ratio with an indication by asterisk if the ratio was significant at the 0.05 level.

EXHIBIT E-10 ANALYSIS OF CLASSROOM EFFECTS WITHIN A SAMPLE OF CENTERS

Number of Classes	Program Length (S, M, L)	Urbanization	Percent Non-White in Sample	Number in Sample	Test	Estimate of Treatment Variance Component	Estimate of Error Variance Component	MS_B/MS_W	F
2	S	Urban	60.0	15	PPVT	-8.94	76.68	12.84/ 76.68	0.17
					PSI	-20.17	198.58	54.49/198.58	0.27
					VSMS	3.86	4.45	32.40/ 4.45	7.06*(1)
					BI	152.58	382.72	1,488.40/382.72	3.89
2	M	Suburban	33.0	12	PPVT	0.38	99.82	102.09/ 99.82	1.02
					PSI	5.59	74.30	108.00/ 74.30	1.45
					VSMS	2.73	6.25	22.69/ 6.25	3.63
					BI	-72.30	445.62	10.09/445.62	0.02
2	M	Urban	7.0	14	PPVT	-3.86	106.74	80.10/106.74	0.75
					PSI	6.96	222.07	270.06/222.07	1.22
					VSMS	6.37	6.27	50.21/ 6.27	8.01*
					BI	54.20	564.39	938.16/564.39	1.66
2	S	Urban	29.0	14	PPVT	-9.82	71.69	2.57/ 71.69	0.04
					PSI	9.16	114.07	178.57/114.07	1.57
					VSMS	-1.01	7.20	0.08/ 7.20	0.01
					BI	29.15	175.36	380.64/175.36	2.17
2	M	Rural	0.0	15	PPVT	-20.17	156.07	9.92/156.07	0.06
					PSI	-35.78	309.16	49.87/309.16	0.16
					VSMS	0.42	4.22	7.23/ 4.22	1.71
					BI	-75.28	633.49	88.01/633.49	0.14
4	L	Urban	100.0	13	PPVT	58.45	153.48	330.59/153.48	2.15
					PSI	62.06	208.89	396.96/208.89	1.90
					VSMS	48.34	14.91	161.40/ 14.91	10.82*
					BI	262.36	776.27	1,571.29/776.27	2.02
4	L	Urban	50.0	12	PPVT	-8.83	80.23	55.69/ 80.23	0.69
					PSI	-28.59	92.96	13.53/ 92.96	0.15
					VSMS	-0.32	16.35	15.47/ 16.35	0.96
					BI	6.65	297.84	316.31/297.84	1.06
4	L	Urban	100.0	13	PPVT	11.73	121.63	159.47/121.63	1.31
					PSI	-38.82	163.21	38.00/163.21	0.23
					VSMS	-3.43	14.83	3.76/ 14.83	0.25
					BI	15.96	213.66	265.13/213.66	1.24
4	S	Urban	100.0	14	PPVT	20.79	81.40	155.64/ 81.40	1.91
					PSI	77.55	118.69	395.64/118.69	3.33
					VSMS	9.69	6.21	40.83/ 6.21	6.57*
					BI	11.68	496.80	436.17/394.44	1.11
4	L	Urban	75.0	12	PPVT	-57.76	194.48	34.04/194.48	0.18
					PSI	6.91	54.68	97.84/ 54.68	1.79
					VSMS	1.97	5.53	10.99/ 5.53	1.99
					BI	-11.41	327.40	10.57/327.40	0.03

Note: (1) An asterisk indicates that the value is significant at the 0.05 level.

It is apparent from Exhibit E-10 that the only significant differences in the two samples of CDC's occurred with the VSMS raw score as the measure of performance. It is true that the sample sizes within CDC's and classes are small indeed. Thus, only relatively large differences were likely to have been detected.

These results suggest that a wide range of factors associated with, or assignable to, different classes within CDC's may have no appreciable effect on the performance or status of the children as measured by these instruments.. Viewed otherwise, the amount of treatment variance is generally small relative to the total center variance ($\hat{\sigma}_A^2 \;/\; \hat{\sigma}_A^2 + \hat{\sigma}_e^2$) .

There are a number of considerations to be borne in mind in evaluating the results shown in Exhibit E-10. The interpretation of the treatment variance, for example, is not unambiguous. No attempt was made here to develop an independent criterion to measure teacher differences or other classroom differences. Also, testers and treatments are confounded, since the same tester generally tested each class in a given CDC. This was not always the case, but in no case did two testers ever test the same children.

There was one case in which the CDC had divided its classes between two buildings. Two testers were assigned to the center. One tester tested children from the two classes in one building; the other tested samples from the three classes in the other building. An analysis of variance of scores for each of the four measures, classified by tester, was made. None of the results of the F-tests was significant at the 0.05 level. For the same center, the F-ratios for the PPVT and the BI were significant at the 0.05 level when analysis of variance was made for classes. The implication is that there was some effect of the variables associated with classes; but not with the testers, contributing to the results for this center. Thus, there is some evidence, at least, that tester effects may not have been a serious source of bias in the data of this study. This point is discussed and analyzed further in Section V.

It is possible that the significant F-ratios for the VSMS measure, when they occurred, were the result of a tester (interviewer)/teacher interaction rather than a teacher/pupil effect. Significant F-ratios for other dependent variables could have resulted from the systematic assignment of children of similar ages to the same class. This apparently was not the case in these samples. However, this would be a source of confounding that should be considered by anyone attempting a more refined study along these lines. The factors contributing to the "within" variance are many--age, sex, race, length of time in program, and so on--since it was not feasible to make analyses controlling for these and other background factors within CDC's.

It is tempting to conclude that the effect of the individual teacher was not very great. However, such a conclusion would have to be based, at the very least, on the assumption that the different teachers, even within a CDC, were attempting to achieve the same goals, or at least the same specifiable and measurable goals. It should also be supported by outside or other evidence that there were substantial differences between teachers on relevant variables. In PRC's data, the range of variation has not been examined in any systematic way.

We cannot draw a final conclusion about teacher effects in Head Start programs from the available data. We can only incline toward acceptance of the hypothesis that the immediate effects are not dramatically large with the classes and children studied.[1]

---

[1] This conclusion is consistent with findings of Coleman, et al, about the relative size of teacher effects in verbal achievement of first graders. (See U.S. Office of Education, OE-38001, Equality of Educational Opportunity, James S. Coleman, et al, 1966; especially Table 3.25.1, p. 317.)

APPENDIX F

THE DRAW-A-PERSON TEST RESULTS

In addition to the four major evaluation measures (PPVT, PSI, BI, and VSMS) a Draw-A-Person (DAP) test was administered to each child. Although the DAP has been widely used as a measure of intelligence and has major advantages in that it is relatively simple and inexpensive to administer and score, little is known about its validity for young disadvantaged children.

PRC's sample of Head Start children provided an opportunity to estimate the validity of the DAP test for young children from low-income families by obtaining DAP test scores and comparing them with PPVT and other scores.

After the completion of the PSI and the PPVT, the tester gave the child a blank sheet of paper (8-1/2 by 11 inches) and said, "Draw a person. Draw the best person you can." If the child asked whether he should draw a man or a woman, the tester replied that the child should draw whichever he wished. In the few instances when the child indicated that he did not know what a "person" was, the tester explained by pointing out familiar examples of persons or people, such as the teacher, a parent, the child himself, etc. Of the 964 tests obtained, 111 were seen as drawings of women, in that the figure had long hair, was obviously clothed in a dress, or was described by the child himself as a female (e.g., "my mommy," "my sister," or " a lady"). Two hundred forty-seven drawings were unrecognizable (Class A), and seven of these were nonresponses in which the child did not put even one mark on the test paper. The remaining 606 drawings were considered to be drawings of men.

The scoring procedures and criteria used were those presented by Harris (1963).[1] Drawings of women were scored according to the

---

[1] Dale B. Harris, Children's Drawings as Measures of Intellectual Maturity. New York: Harcourt, Brace and World, Inc., 1963

Draw-A-Woman criteria, and drawings of men were scored according to the Draw-A-Man criteria. Blank or unrecognizable drawings were marked Class A and received a raw score of 0 or 1, in accordance with Harris' criteria. As noted above, only 7 of the 247 Class A drawings were blank.

Scoring of the drawings was done by a team of four scorers.[1] It was decided that each of a sample of drawings should be scored independently by each of the four scorers until sufficiently high interscorer reliabilities were obtained. Exhibit F-1 indicates the interscorer reliability coefficients that were obtained on a sample of drawings before the scorers began dividing the scoring task so that each drawing would be scored only once. Note in Exhibit F-1 that correlations were computed for Class B drawings only, as well as for Class A and Class B combined. Since Class B drawings always had a score of 2 or more and Class A drawings always had either a score of 0 or 1, it was felt that a correlation between Class B drawings would yield a more rigorous check on the scorers' agreement. Exhibit F-1 also includes an interscorer correlation coefficient which was obtained on a small sample of drawings after about half the drawings had been scored. It can be seen that all interscorer correlation coefficients were 0.89 or above.

Exhibits F-2, F-3, and F-4 present the distributions of DAP raw scores by children's sex, race, and age, for Short-term, Medium-term, and Long-term centers. These exhibits follow the same format as Exhibits 19A through 22C in subsection IV.B.1.a.

Exhibit F-5 presents intertest product-moment correlation coefficients among the DAP raw score (RS), DAP IQ, PPVT raw score (RS), PPVT IQ, and PSI raw score (RS). Parts A through F of the exhibit present

---

[1] Scorers were: Miss Marilyn Danner, a psychology major who had extensive experience in scoring drawings for a research project under the direction of Dr. Dale Harris and Dr. Elizabeth Kirchner at Pennsylvania State University; Dr. Lois-ellin Datta, psychologist in the Personality Section at National Institute of Mental Health; Miss Carol J. Kline, doctoral candidate in clinical psychology at the University of Maryland; and Dr. Ruth Ann O'Keefe, PRC educational psychologist.

EXHIBIT F-1 INTERSCORER RELIABILITY FOR PERSON DRAWINGS

Scorer	All Drawings in Small Sample(1)	Class B Drawings Cnly(2)
B - C	0.979 (N = 42)	0.963 (N = 27)
B - D	0.989 (N = 50)	0.984 (N = 32)
D - A	0.989 (N = 42)	0.982 (N = 25)
D - A(3)	0.976 (N = 15)	
C - A(3)	0.894 (N = 14)	

Notes: (1) Class A and Class B

(2) Class B drawings are recognizable drawings which obtain a raw score of 2 or more. Class A drawings are blank (i.e., non-drawings) or unrecognizable, and can obtain raw scores of 0 or 1. For this category, Class A drawings were eliminated from the computations.

(3) These correlation coefficients were obtained after approximately one-half of the total number of drawings had been scored.

EXHIBIT F-2 DISTRIBUTION OF DAP RAW SCORES FOR CHILDREN IN SHORT-TERM CENTERS, BY SEX, RACE, AND AGE

	Male															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	01 05 05	00 07	01 01 03 04 07 13	02 03 04 05 05 06 07 07 08 15	00 02 05 06 07 08 08 10 11 12 25	00 01 01 02 07 08 09 11 12 13 14 16 18 18 21 31	00 00 01 01 07 12 13 20 22	05	00 00 00 00 01 04 04 05	00 00 00 00 00 01 01 01 01 01 01 01 01 01 01 04 05 07 08 09 10 11 13	00 00 00 01 01 01 01 01 01 01 01 01 01 03 03 03 04 05 06 06 06 07 08 08 09 12 12 14 18	00 00 01 01 01 01 02 03 04 04 05 05 05 05 06 06 06 07 08 09 10 11 12 14 14 15 15 20	04 04 05 07 08 10	08 16 17	05 09
Total																
N	-	3	2	6	10	11	16	9	1	8	23	29	28	6	3	2
Mean	-	3.67	3.50	4.83	6.20	8.55	11.38	8.44	5.00	1.75	3.35	4.62	6.79	6.33	13.62	7.00
Median	-	5.00	3.50	3.50	7.50	8.00	11.50	7.00	5.00	0.50	1.00	3.00	5.50	6.00	16.00	7.00
Eligible																
N	-	3	2	6	10	11	16	9	1	8	23	29	28	6	3	2
Mean	-	3.67	3.50	4.83	6.20	8.55	11.38	8.44	5.00	1.75	3.35	4.62	6.79	6.33	13.62	7.00
Median	-	5.00	3.50	3.50	7.50	8.00	11.50	7.00	5.00	0.50	1.00	3.00	5.50	6.00	16.00	7.00

EXHIBIT F-2 (Continued)

	Female															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	00	01	00	00	00	01	01	06	-	01	00	00	00	08	04	00
		01	01	01	01	07	05	10		01	01	01	01	11	05	06
		01	01	01	01	09	05	12		01	01	01	01	15	09	07
			06	01	07	09	09	17		05	01	01	01	16	09	08
				01	08	10	10	18		07	01	01	05		10	16
				03	08	13	12	18		10	01	01	05		12	
				04	09	13	13	19			03	01	05		16	
				06	10	16	14	21			03	02	05		17	
				07	11	20	14	22			07	02	06			
				08	12	20	16	34			07	03	07			
				09	14		17	36			08	03	07			
				14	15		17				09	03	08			
				15	22		23				09	03	08			
				17	23		28				09	04	08			
											10	04	09			
											10	04	10			
											11	04	10			
												05	12			
												05	16			
												05	18			
												05	19			
												05	19			
												06				
												06				
												06				
												06				
												06				
												07				
												07				
												07				
												08				
												10				
												11				
												14				
Total																
N	1	3	4	14	14	10	14	11	-	6	17	34	22	4	8	5
Mean	0.00	1.00	2.00	6.21	10.07	11.80	13.14	19.36	-	4.16	5.35	4.61	8.18	12.75	10.25	7.40
Median	0.00	1.00	1.00	5.00	9.50	12.50	13.50	18.00	-	3.00	7.00	4.50	7.50	13.50	9.50	7.00
Eligible																
N	1	3	4	14	14	10	14	11	-	6	17	34	22	4	8	5
Mean	0.00	1.00	2.00	6.21	10.07	11.80	13.14	19.36	-	4.16	5.35	4.61	8.18	12.75	10.25	7.40
Median	0.00	1.00	1.00	5.00	9.50	12.50	13.50	18.00	-	3.00	7.00	4.50	7.50	13.50	9.50	7.00

EXHIBIT F-3 DISTRIBUTION OF DAP RAW SCORES FOR CHILDREN IN MEDIUM-TERM CENTERS, BY SEX, RACE, AND AGE

	Male															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	02	00 01 01*(1) 03 03 04 10*	00 01 01 01 01 01 01 01 01* 02* 03 03 04 04 04 04 05 05 05 05 05* 06 06 06 06 06 07 07 08 08 09* 09* 10 11 12 14 15	00 01 01 01 01 02 02* 04 05 05 05 05 05 05 06* 06 06 06 07 07 07* 07 08 08 08* 08 08 08 09* 09 09 09 09 12 15* 18* 19*	01* 01 03* 03 04 04 04 04 05 06* 06 06 06 07 08 08 09* 09 09 10 10* 12 13 14	01 06 06 06 07 07 08 08 09 10 11 11 11 11 12 12 12 13 14 15 15 17 17 20 24 24	06 06 08 09* 10* 14 16 19	01 03*	00 00 00 00 01 01	00 00 00* 00 00 00 00 01 02 05 06* 08*	00* 01 01* 03* 04* 04* 05 06 06 06* 06 06 07 08 12 13 14	01 01 04 04 05 06 06 07* 08 10* 12* 12*	04	06	05
Total																
N	-	1	7	37	38	24	26	8	2	6	12	17	12	1	1	1
Mean	-	2.00	3.14	5.32	6.76	6.75	11.81	11.00	2.00	0.33	1.83	6.00	6.33	4.00	6.00	5.00
Median	-	2.00	3.00	5.00	6.50	6.00	11.00	9.50	2.00	0.00	0.00	6.00	6.00	4.00	6.00	5.00
Eligible																
N	-	1	5	32	30	19	26	6	1	6	9	11	8	1	1	1
Mean	-	2.00	2.20	5.34	5.77	7.00	11.81	11.67	1.00	0.33	0.89	7.64	4.38	4.00	6.00	5.00
Median	-	2.00	3.00	5.00	6.00	6.00	11.00	8.50	1.00	0.00	0.00	6.00	4.50	4.00	6.00	5.00

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 17 weeks.

EXHIBIT F-3 (Continued)

	Female															
	White								Non-White							
	Age Range								Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	01 01 02 04	01 01 04 06 06 07 13	00*(1) 00* 01 01* 01 01 01 01 01 01 02 04 06* 06* 06 06 07 07 07 08 08 08 08 09 11 13 13 14* 14 15	01 01 01 01 01 01 01 01* 04 05* 05* 05 05 06 06 06 07 07 07 07* 08* 08* 09 09 10 10* 10 12 13 22	00 01 01 05 05 06 08 09 09 10 13 13 14 15 17 17 17 18 24 34	01 01 06 06 10 11 11 12 13 13 14 14 15 15 17 17 22	11 12 15 22 24 25 26	00*	00 01 01 01 01 05 07	00 01* 01 04 09	00 01 03 06* 06 06 07 08* 08 08 09 10 12	00* 01* 01* 01 01 02* 08 09 10* 17 18	00 08 09 14 15	27	3 16
Total																
N	-	4	7	30	30	20	17	7	1	7	5	13	11	5	1	2
Mean	-	2.00	5.43	6.00	6.30	11.80	11.65	19.29	0.00	2.29	3.00	6.46	6.18	9.20	27.00	9.50
Median	-	1.50	6.00	6.00	6.00	11.50	13.00	22.00	0.00	1.00	1.00	7.00	2.00	9.00	27.00	9.50
Eligible																
N	-	4	7	24	23	20	17	7	0	7	4	11	6	5	1	2
Mean	-	2.00	5.43	6.38	6.30	11.80	11.65	19.29	0.00	2.29	3.50	6.36	9.00	9.20	27.00	9.50
Median	-	1.50	6.00	7.00	6.00	11.50	13.00	22.00	0.00	1.00	2.50	7.00	8.50	9.00	27.00	9.50

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 17 weeks.

EXHIBIT F-4 DISTRIBUTION OF DAP RAW SCORES FOR CHILDREN IN LONG-TERM CENTERS, BY SEX, RACE, AND AGE

	Male							
	White							
	Age Range							
	3-0 3-5	**3-6 3-11**	**4-0 4-5**	**4-6 4-11**	**5-0 5-5**	**5-6 5-11**	**6-0 6-5**	**6-6 6-11**
	01*(1)	-	01	01 01 04 04 05* 05 06 07 07* 08 11	00 01* 01 01* 03 06 12 14 15	04 08*	-	-
Total								
N	1	-	1	11	9	2	-	-
Mean	1.00	-	1.00	5.36	5.89	6.00	-	-
Median	1.00	-	1.00	5.00	3.00	6.00	-	-
Eligible								
N	0	-	1	9	7	1	-	-
Mean	0.00	-	1.00	5.22	7.29	4.00	-	-
Median	0.00	-	1.00	5.00	6.00	4.00	-	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT F-4 (Continued)

	Male							
	Non-White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	02	00	01	01	01	01	13	-
	02*(1)	00	01	01*	01	05*		
		00	01	01	01*	06		
		01	01	01	01*	08		
		01	01	01	01			
		01	01	01	03			
		01*	01*	01	04*			
		01*	03	01	05			
		01*	04*	01	05			
		01	05	03*	05			
		01*	05	03	05			
		03*	06	03	06			
		05	06	03	06			
		09	08*	03	06			
		09*	12	03	06			
		09		04	06			
				04	07			
				04	07*			
				04	07*			
				05*	08			
				05	08			
				05	09			
				05	10			
				05	11			
				05	11			
				05	12			
				05*	12			
				06	12			
				06	13			
				06*	18			
				06	21			
				06				
				06				
				06				
				06				
				07*				
				07				
				07				
				07				
				07				
				08*				
				08*				
				08*				
				10*				
				10				
				10				
				10*				
				11				
				12				
				14				
				15				
				16				
				24				
Total								
N	2	16	15	53	31	4	1	-
Mean	2.00	2.69	3.73	6.08	7.35	5.00	13.00	-
Median	2.00	1.00	3.00	5.00	6.00	5.50	13.00	-
Eligible								
N	1	10	12	42	26	3	1	-
Mean	2.00	2.70	3.58	5.98	8.00	5.00	13.00	-
Median	2.00	1.00	2.00	5.00	6.50	6.00	13.00	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT F-4 (Continued)

	Female							
	White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	-	01 06	03 08	01 07 13	01*(1) 01 01 09 14 15 17*	12	-	-
Total								
N	-	2	2	3	7	1	-	-
Mean	-	3.50	5.50	7.00	8.29	12.00	-	-
Median	-	3.50	5.50	7.00	9.00	12.00	-	-
Eligible								
N	-	2	2	3	5	1	-	-
Mean	-	3.50	5.50	7.00	8.00	12.00	-	-
Median	-	3.50	5.50	7.00	9.00	12.00	-	-

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT F-4 (Continued)

	Female							
	Non-White							
	Age Range							
	3-0 3-5	3-6 3-11	4-0 4-5	4-6 4-11	5-0 5-5	5-6 5-11	6-0 6-5	6-6 6-11
	01	00*(1)	00	00	00	01	15	-
	02	01	01*	01	01	09		
	04	01*	01*	01	01*	11		
		01	01	01	01	14		
		01	02*	01	01			
		03	02	01	02			
			04	01*	03			
			04	01*	03			
			04	01*	04			
			05*	01	04			
			06*	01	04			
			06	02*	05			
			07*	03*	05*			
			07*	03	05			
			07*	03*	05			
			08	04	06*			
			13	05*	06			
			14	05	06*			
			14*	06	07			
				06	08			
				06*	08*			
				06*	08			
				06	08			
				07	09			
				07	10			
				08*	10			
				08	10*			
				08*	10			
				08	11			
				09	11			
				09	11			
				09	11			
				10*	12*			
				10	13			
				10	13			
				10*	13*			
				11	13*			
				11	13			
				12	14			
				12	15			
				12	15			
				13*	17*			
				13*	20*			
				13	22			
				13	24			
				15				
				18*				
				18*				
Total								
N	3	6	19	48	45	4	1	
Mean	2.33	1.17	5.58	7.06	8.93	8.75	15.00	
Median	2.00	1.00	5.00	7.00	8.00	10.00	15.00	
Eligible.								
N	3	4	10	31	34	4	1	
Mean	2.33	1.50	5.60	6.87	8.56	8.75	15.00	
Median	2.00	1.00	4.00	7.00	8.00	10.00	15.00	

Note: (1) An asterisk denotes an ineligible child, who was in the program for less than 25 weeks.

EXHIBIT F-5 INTERTEST CORRELATIONS FOR TOTAL AND SUBGROUPS: PPVT RAW SCORE, PPVT IQ, DAP RAW SCORE, DAP IQ, AND PSI RAW SCORE

A.	All Children (N = 956)[1]				
	PPVT RS	PSI RS	PPVT IQ	DAP RS	DAP IQ
PPVT RS	1.00	0.73	0.84	0.46	0.31
PSI RS.		1.00	0.53	0.56	0.36
PPVT IQ			1.00	0.25	0.29
DAP RS				1.00	0.85
DAP IQ					1.00

B.	Female, White (N = 199)				
	PPVT RS	PSI RS	PPVT IQ	DAP RS	DAP IQ
PPVT RS	1.00	0.78	0.79	0.50	0.40
PSI RS		1.00	0.59	0.58	0.43
PPVT IQ			1.00	0.23	0.32
DAP RS				1.00	0.88
DAP IQ					1.00

C.	Female, Non-White (N = 263)				
	PPVT RS	PSI RS	PPVT IQ	DAP RS	DAP IQ
PPVT RS	1.00	0.65	0.83	0.44	0.28
PSI RS		1.00	0.40	0.54	0.30
PPVT IQ			1.00	0.24	0.30
DAP RS				1.00	0.84
DAP IQ					1.00

Note: (1) This includes all children for whom scorable results on all five measures (PPVT, PSI, DAP, VSMS, and BI) were obtained.

EXHIBIT F-5 (Continued)

D.	Male, White (N = 188)(1)				
	PPVT RS	PSI RS	PPVT IQ	DAP RS	DAP IQ
PPVT RS	1.00	0.69	0.88	0.40	0.23
PSI RS		1.00	0.07	0.51	0.13
PPVT IQ			1.00	0.20	0.19
DAP RS				1.00	0.89
DAP IQ					1.00

E.	Male, Non-White (N = 274)				
	PPVT RS	PSI RS	PPVT IQ	DAP RS	DAP IQ
PPVT RS	1.00	0.72	0.87	0.51	0.36
PSI RS		1.00	0.55	0.56	0.39
PPVT IQ			1.00	0.37	0.36
DAP RS				1.00	0.88
DAP IQ					1.00

F.	Children Tested in Spanish (N = 65)				
	PPVT RS	PSI RS	PPVT IQ	DAP RS	DAP IQ
PPVT RS	1.00	0.72	0.86	0.53	0.47
PSI RS		1.00	0.52	0.60	0.54
PPVT IQ			1.00	0.41	0.44
DAP RS				1.00	0.96
DAP IQ					1.00

Note: (1) Computations include only white boys tested in English and omit white boys tested in Spanish.

the intertest correlations for (A) all children, and subgroups of (B) white females only, (C) non-white females only, (D) English-speaking white males only, (E) non-white males only, and (F) children tested in Spanish.

Exhibit F-6 presents, for 3-, 4-, 5-, and 6-year-old children separately, intertest correlation coefficients among the DAP and PPVT raw and IQ scores and PSI raw scores.

Exhibits F-5 and F-6 show that the correlations between DAP and PPVT raw scores for all subgroups compare favorably with the 0.39 obtained in a large sample of children in the summer 1965 Head Start program.[1] The DAP-PPVT IQ correlations, ranging from .19 to .44, were generally lower than the raw score correlations.[2]

Exhibit F-7 presents unadjusted raw score means, IQ's, and SQ's derived from scores of the PPVT, DAP, and VSMS, as well as standard deviations of all means, for the 956 children who completed all five measures (PPVT, PSI, DAP, VSMS, and BI). The scores and standard deviations are organized by sex and race of children tested.

The exhibit shows that all mean IQ estimates (PPVT and DAP) for all groups were lower than means of normative samples. (The PPVT IQ means ranged from 79.31 to 87.27, and the DAP IQ means from 77.22 to 78.71.) The low PPVT scores are consistent with the verbal handicap frequently reported for children from low-income families, but the finding that DAP IQ's for all groups were even lower than PPVT IQ's was unexpected. For example, five groups of children (sample sizes ranging from 17 to 500) in the summer 1965 Head Start project were tested on both the PPVT and the DAP. In all five cases,

---

[1] Planning Research Corporation, PRC R-795, Results of the Summer 1965 Project Head Start, H.R. Cort, Jr., et al, May 1966, Vol. I, p. IV-98. (Report prepared for the Office of Economic Opportunity under Contract OEO-753.)

[2] Dr. Lois-ellin Datta was responsible for processing and analysis of DAP scores and intertest correlations. She is preparing three separate papers based on DAP results from the PRC sample. See footnote 2, page F-17.)

EXHIBIT F-6 INTERTEST CORRELATIONS FOR THREE-, FOUR-, FIVE-, AND SIX-YEAR-OLD PPVT RAW SCORE, PPVT IQ, DAP RAW SCORE, DAP IQ, AND PSI RAW SCORE

	Age	3	4	5	6	Total
Tests	N	72	407	346	131	956
PPVT RS/PSI		0.48	0.70	0.63	0.71	0.73
PPVT RS/PPVT IQ		0.89	0.95	0.91	0.98	0.84
PPVT RS/DAP RS		0.39	0.35	0.31	0.39	0.46
PPVT RS/DAP IQ		0.37	0.35	0.34	0.42	0.31
PSI/PPVT IQ		0.34	0.64	0.60	0.68	0.53
PSI/DAP RS		0.39	0.41	0.45	0.44	0.56
PSI/DAP IQ		0.38	0.40	0.45	0.46	0.36
PPVT IQ/DAP RS		0.24	0.29	0.30	0.34	0.25
PPVT IQ/DAP IQ		0.22	0.28	0.33	0.39	0.29
DAP RS/DAP IQ		0.99	0.98	0.98	0.98	0.85

EXHIBIT F-1 RAW SCORE IQ AND SQ MEANS AND STANDARD DEVIATIONS

		PPVT RS	PSI RS	PPVT IQ	DAP RS	DAP IQ	VSMS RS	VSMS SQ	BI RS
All subjects N = 956⁽¹⁾	Mean	39.30	46.82	82.02	6.95	77.22	56.91	106.33	142.54
	SD	11.48	21.82	19.44	5.69	16.01	6.64	19.95	23.41
Male, white N = 220	Mean	45.22	59.50	87.27	7.35	77.03	58.13	103.91	142.71
	SD	11.01	36.35	20.62	5.39	15.21	6.56	18.93	22.28
Male, non-white N = 274	Mean	36.27	40.39	80.65	5.19	76.05	55.49	106.89	139.09
	SD	11.32	13.26	18.76	4.55	16.68	6.09	21.14	24.99
Female, white N = 199	Mean	41.35	50.25	81.66	9.07	78.71	57.38	102.86	142.12
	SD	11.47	15.64	20.98	7.06	16.91	6.12	19.28	23.22
Female, non-white N = 263	Mean	36.57	43.47	79.31	6.86	77.46	57.02	110.43	146.31
	SD	9.57	12.70	17.63	5.23	15.22	6.48	19.33	22.30

Note: (1) This figure is less than the actual total of 964, due to blanks on the DAP tests.

the DAP IQ estimate was higher than the PPVT IQ estimate (mean DAP IQ range was 76 to 111, and mean PPVT IQ range was 68 to 91).[1]

Dr. Lois-ellin Datta has undertaken three separate studies involving the DAP test results obtained from the PRC samples; one of these examines in some detail the findings shown in Exhibits F-5, F-6, and F-7.[2] The other two studies, co-authored with Ann Drake of the National Institutes of Health, concern "Objective Sexual Differentiation in the Drawings of Preschool Children" and "The Smiling Faces: A Comparison of the Effect of Drawings of Children from Middle Class and Very Low Income Families." The former has been submitted to the Journal of Consulting Psychology, and is available in mimeographed form from Dr. Datta.[2] The latter, which will incorporate a study of motivational factors relating to performance on the DAP, is currently in preparation.

---

[1] Ibid., H.R. Cort, et al, Vol. II, p. E-2.

[2] Datta, Lois-ellin, "The Draw-A-Person Test as a Measure of Intelligence in Preschool Children from Very Low Income Families," Journal of Consulting Psychology, to be published in April or June 1968. (Mimeographed copies are currently available from: Dr. Lois-ellin Datta, 2N256, Bldg. 10, National Institutes of Health, Bethesda, Maryland 20014.)